# Analytic and Geometric Aspects of Gauge Theory

**Aleksander Doan**

Modern physics paints a picture of reality radically different from our everyday experience. In that picture, all physical interactions propagate through certain fields, which permeate space and time, and it is those fields, and not particles or forces, that are the most fundamental building blocks of nature. At the core of this theory lie two facts, which connect the physics of fields to the problems of pure mathematics. First, the dynamics of fields is governed by systems of partial differential equations. Second, these equations exhibit special symmetries which do not come from the symmetries of space-time. Rather, they reflect the freedom in the choice of a mathematical description of fields, which physicists refer to as a *gauge*. Thus, *gauge theory* is the study of fields with such symmetries. While physics is interested chiefly in the *quantum* theory of fields, it turns out that already the classical, non-quantum theory has an immensely rich mathematical structure, whose study over the last forty years has led to many discoveries at the intersection of analysis, geometry, and topology.

## Maxwell's Equations and Topology

The prototypical example of a theory with gauge symmetry is Maxwell's theory of electromagnetism. If we think of space-time as a four-dimensional manifold $M$, then the electric and magnetic fields are represented by vector fields on $M$. There is a convenient way of combining them into a single mathematical entity: the electromagnetic field $F$, which is a differential two-form on $M$, that is, an expression of the form $F = \sum_{ij} F_{ij} dx_i dx_j$, where $x_i$ are coordinates and $F_{ij}$ functions on $M$. There are two linear differential operators acting on the spaces of forms, the exterior derivative $d$ and its adjoint $d^*$, corresponding to the curl and div operators known from vector calculus. According to Maxwell's theory, $F$ obeys a system of partial differential equations, which in the simplest case is

$$dF = 0 \quad \text{and} \quad d^*F = 0.$$

These equations make sense on any manifold $M$ equipped with a metric (which in physics would have indefinite signature, meaning that distances can be negative, but here we take it to be positive, like in Euclidean geometry). Given such $M$, we can ask whether there are any solutions to Maxwell's equations and if they tell us something about the geometry of $M$.

The first observation is that these two first order equations can be a reduced to a single second order equation. The exterior derivative satisfies $dd = 0$, so if we find a differential one-form $A$ satisfying $F = dA$ (which we can always do, at least in a neighborhood of every point of $M$, by the fundamental theorem of calculus), then the first equation is automatically satisfied and we are left with

$$d^*dA = 0. \qquad (\dagger)$$

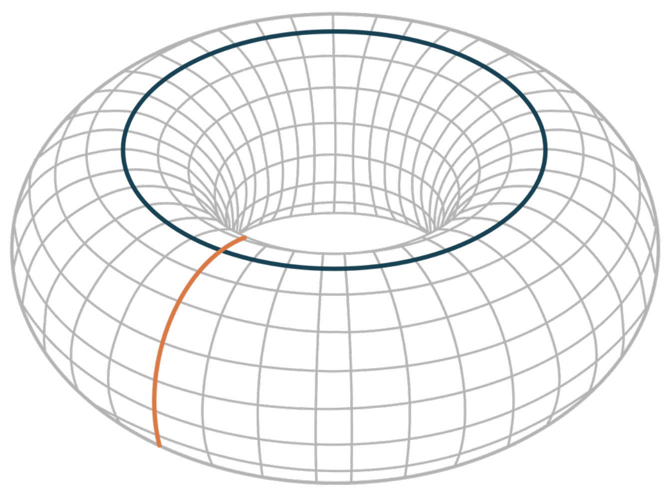
This simplification, however, comes at the price of ambiguity, as there are infinitely many choices of $A$ satisfying $F = dA$, and it is $F$, not $A$, that has a physical meaning. For example, for any function $f\colon M \to \mathbb{R}$ we can replace

$$A \mapsto A + df, \qquad (\ddagger)$$

which does not change $dA$ since $ddf = 0$. Thus, we are interested in the space of all $A$'s solving (†) up to transformation (‡), which we call a gauge symmetry. In practice, it is hard to directly study system (†), as it is underdetermined: it has more unknown functions than equations. One way to fix this issue is to get rid of gauge symmetries by finding $A$ which obeys the additional *gauge fixing* condition $d^*A = 0$, so that (†) becomes equivalent to the Laplace equation

$$\Delta A = (d^*d + dd^*)A = 0,$$

which has much better analytical properties. In particular, if $M$ is compact, an old theorem of Hodge tells us that the space of solutions can be identified with the cohomology group $H^1(M)$, a classical topological invariant of $M$. We have reached a surprising conclusion that Maxwell's theory encodes some topological information about the underlying manifold. For example, if $M$ is a torus, $H^1(M)$ is two-dimensional, corresponding to the fact that the electromagnetic field on $M$ can wind around the torus in two different directions.





Horizontal and vertical lines represent two different ways in which the electromagnetic field can wrap around a torus.

It is exactly the winding behavior of fields that allows us to interpret (‡) as a symmetry. From a rather abstract point of view, the one-form $A$ can be interpreted as a geometric structure called *connection* on the manifold $M \times \mathbb{R}^2$. Here we think of $M \times \mathbb{R}^2$ as having a copy of the plane $\mathbb{R}^2$ for every point of $M$, and $A$ tells us how these planes rotate as we travel in $M$. In this framework, $F$ is the *curvature* of the connection, while transformation (‡) corresponds to changing coordinates on $M \times \mathbb{R}^2$ by rotating the plane over $x \in M$ by the angle $f(x)$ modulo $2\pi$, for each $x$. While $A$ changes under this transformation, the geometry of the entire setup remains the same, and the invariance of (†) under (‡) is a manifestation of this fact. In other words, gauge symmetries of Maxwell's theory are related to $SO(2)$, the group of rotations of the plane. An important generalization of this idea is to replace $M \times \mathbb{R}^2$ by an arbitrary *bundle* of planes over $M$, for which the copies of $\mathbb{R}^2$ can twist in a topologically nontrivial way, similarly to how the Möbius band twists when we travel around it.

## Yang–Mills Theory

In the 1950s, the physicists Yang and Mills proposed a theory of nuclear forces based on partial differential equations generalizing Maxwell's equations. In this theory, the group $SO(2)$ is replaced by a general matrix group, such as the group of $n$-dimensional isometries $SO(n)$ or the unitary groups $U(n)$, $SU(n)$. What seems like a merely technical generalization leads to many completely new features. Unlike $SO(2) = U(1)$, higher rank matrix groups are non-commutative. This results in nonlinear field equations, as $A$ is now a connection represented by a matrix of one-forms, and, after gauge fixing, the Yang–Mills equations have the form

$$\Delta A + f(A, dA) = 0,$$

where $f(A, dA)$ consists of quadratic and cubic combinations of commutators of $A$ and $dA$. In physics, these terms correspond to the interactions of the field with itself. Much is known about the analysis of these equations, largely due to the pioneering work of Uhlenbeck, who described a mechanism by which the energy of a sequence of solutions on a manifold $M$ can become infinitely concentrated along a manifold of lower dimension contained in $M$.

Another new feature is that, unlike in the case of electromagnetism, the spaces of Yang–Mills connections up to gauge symmetry, so-called *moduli spaces*, can have very complicated geometry. The study of these spaces is particularly fruitful for manifolds of dimension four, which admit a special class of solutions to the Yang–Mills equations called *instantons*, characterized by having minimal energy. For instance, all instantons on $M = \mathbb{R}^4$ can be constructed by algebraic methods, and their moduli spaces provide examples of *hyperkähler manifolds*: rare objects studied in algebraic and differential geometry, whose structure is related to the algebra of quaternions.

## Invariants of Manifolds

In the 1980s, Donaldson proved a number of astonishing theorems about four-dimensional manifolds by relating the geometry of the moduli spaces of instantons on a four-manifold $M$ to the topology of $M$. One of these theorems states that there exist topological four-manifolds which carry infinitely many inequivalent smooth structures, in contrast with higher dimensions where this is known to be impossible. By a smooth structure we mean here a way of covering the manifold with coordinate charts such that all coordinate changes are differentiable to any order. Similar techniques show that infinitely many (in some sense, most) topological four-manifolds do not carry a smooth structure at all. It is striking that all known proofs of these topological results rely heavily on the analysis of the partial differential equations of gauge theory.

What is now known as Donaldson theory opened a new era in the study of four-manifolds. Shortly afterwards, these methods were extended to solve long-standing problems about three-manifolds, knots, and embeddings of surfaces. By now, gauge theory is one of the standard tools at the disposal of every low-dimensional topologist. However, a number of important questions remain open. For instance, it is not clear how much of low-dimensional topology is seen by gauge-theoretic methods. Instantons, as well as solutions to closely related equations introduced by the physicists Seiberg and Witten, lead to invariants of smooth four-manifolds, called Donaldson and Seiberg–Witten invariants, which can be used to tell apart inequivalent smooth structures. But it is still unknown whether we can find smooth structures whose invariants agree but which are nevertheless distinct. Complex geometry provides potential examples, but new tools will have to be introduced in order to actually distinguish them. There are similar questions for the corresponding invariants of three-manifolds and knots.

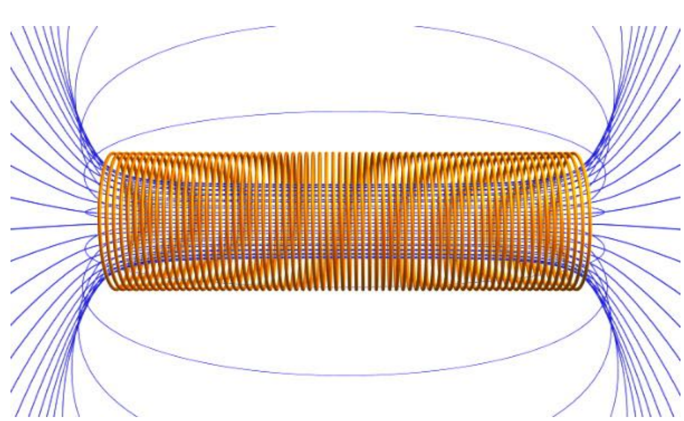


Figure by PerOX, Public domain, via Wikimedia Commons / Color modified

Electric current generating magnetic field, in agreement with Maxwell's equations.

A related question is whether Donaldson and Seiberg–Witten invariants can be refined, exploiting the rich structure of gauge-theoretic equations. Methods of homotopy theory provide one such refinement. This is familiar from algebraic topology: a map between topological spaces $f\colon X \to Y$ induces a map on the homology groups $f_*\colon H_*(X) \to H_*(Y)$, which is useful for studying the topology of $X$ and $Y$. However, more detailed information is encoded in the homotopy class of $f$. Gauge-theoretic invariants are analogous to homological information; in this context, $X$ and $Y$ are infinite-dimensional spaces and $f$ is given by the field equations. Bauer and Furuta introduced homotopical refinements of these invariants, which has proved incredibly useful. A recent application is Manolescu's resolution of a long-standing conjecture about triangulations of manifolds. For more on this topic, we refer to the article on the Floer Homotopy Theory program in this issue of the MSRI Emissary.

Other foundational open questions concern the structure of gauge-theoretic invariants. The most important of them is the simple type conjecture, which asserts that all information encoded in these invariants is, in some sense, finite. There is a related conjecture of Witten, now close to being proved, according to which Donaldson and Seiberg–Witten invariants are related by an intricate, but explicit formula predicted by string theory. An analogous problem for the invariants of three-manifolds is completely open, as their algebraic structure is much more complicated than that of the four-dimensional invariants. The situation is even less understood for the homotopical refinements of these invariants, as so far only Seiberg–Witten theory, and not Donaldson theory, has been generalized in this way. Constructing a homotopical refinement of Donaldson invariants remains a fascinating but technically challenging possibility.

## Geometry of Moduli Spaces

Topological applications of Donaldson theory rely on detailed study of instanton moduli spaces. It turns out that the geometry and topology of these spaces is rather interesting in its own right. We have already mentioned that some of them carry a hyperkähler structure. In fact, all known compact hyperkähler manifolds can be built from some basic examples (tori, K3 surfaces) and instanton moduli spaces. Interesting non-compact examples can be constructed by studying the instanton equation on four-manifolds of the form $M = N \times \mathbb{R}^d$, where $d = 1, 2, 3$ and $N$ is a manifold of dimension $4 - d$. By looking at solutions which are invariant in the $\mathbb{R}^d$ direction, we obtain a system of differential equations on $N$.

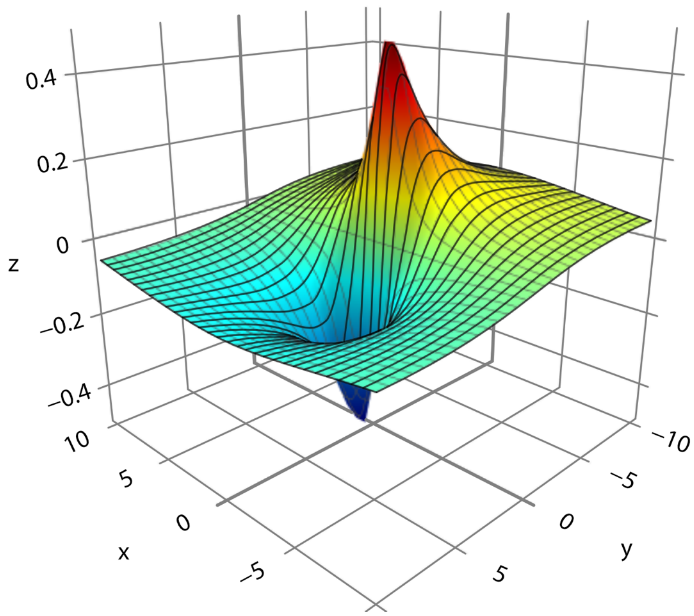
For $d = 1$ and $N$ three-dimensional, these are the Bogomolny monopole equations. Their moduli spaces of solutions on $N = \mathbb{R}^3$ are non-compact hyperkähler manifolds, which appear naturally in the study of certain quantum field theories in physics. This has led to a number of interesting questions about their geometry. For example, the S-duality conjecture makes a precise prediction about the $L^2$ cohomology, a geometric invariant of non-compact manifolds, of these spaces. This conjecture is closely tied to the problem of understanding the asymptotic geometry of these spaces at infinity, a difficult task since the hyperkähler structure is given by an abstract construction and cannot be written explicitly.

A better understood case is $d = 2$, with $N$ being a surface. The instanton equation in this case leads to the Hitchin equations, which turn out to be connected in a fascinating way to many areas of mathematics: representation theory, algebraic geometry, mirror symmetry, and the geometric Langlands program. These connections are discussed in the Fall 2019 issue of the Emissary. We only remark here that the moduli spaces of solutions are again hyperkähler and non-compact, and there is a tantalizing conjecture, originating from string theory, which provides a rather explicit description of the hyperkähler structure at infinity. If proved, it might lead to a geometric way of compactifying the Hitchin moduli spaces, and therefore to new examples of compact hyperkähler manifolds. The key step in proving this conjecture is to describe the limiting behavior of solutions to Hitchin's equations as we travel far in the moduli space, a problem which has already stimulated many exciting developments in geometric analysis.

## Uncharted Territories in Low Dimensions

We have so far discussed various applications of the instanton and Seiberg–Witten equations to the study of manifolds of dimensions four, three, and two. Recent years have witnessed a growing interest in a multitude of other, less understood gauge theories.

On the one hand, there are numerous generalizations of the Seiberg–Witten equations. In fact, there is one for every compact Lie group $G$ and a representation of $G$ on a quaternionic vector space $V$. As before, the group corresponds to gauge symmetries, while $V$ is the space of values of an additional field, called the *Higgs field* by virtue of its relation to the famous Higgs particle. This is familiar from physics, where fields carrying different interactions transform differently under gauge symmetries. Donaldson theory corresponds to $G = SU(2)$ and $V = \{0\}$, while Seiberg–Witten theory to $G = U(1)$ and $V = \mathbb{C}^2$. Other examples include the Kapustin–Witten equations, which conjecturally lead to new topological invariants of three–manifolds and knots, and the Vafa–Witten equations, which have links to algebraic geometry and modular forms. There has been recently a surge in the study of these and related equations, initiated by deep work of Taubes, who showed that their solutions can degenerate in a way unknown from other problems in analysis and geometry. Understanding such degenerations will require developing completely new analytic tools, but it is likely to lead to further applications of gauge theory to topology.






Instanton with energy highly concentrated at a point.

## Towards Higher Dimensions

On the other hand, many attractive features of Yang–Mills theory on four-manifolds generalize to certain manifolds of higher dimensions. Here we are not interested in the classification of manifolds up to smooth equivalence, a problem which can be solved using classical tools of algebraic topology, but rather in understanding a particular class of geometric structures, called *special holonomy metrics*. These structures arise naturally from string theory, where they are related to supersymmetry, but they are also of independent interest to geometers, as they are examples of solutions to Einstein's equations.

From the viewpoint of gauge theory, manifolds with such metrics are interesting because they admit special solutions of the Yang–Mills equations, similar to instantons in dimension four. It is natural to ask whether we can extract invariants of special holonomy manifolds from instanton moduli spaces, mimicking the four-dimensional story. Such invariants would be helpful in classifying the existing millions of examples of special holonomy manifolds, whose geometry is not well understood. This is a fascinating proposal, which in recent years has been the subject of intense research. Like in the low-dimensional situation, here as well the main difficulty lies in understanding degenerations of solutions of the equations. Rather surprisingly, the low- and high-dimensional theories seem to be intimately connected. Based on Uhlenbeck's work, mentioned earlier, it has been conjectured that a sequence of higher-dimensional instantons on a special holonomy manifold $X$ can concentrate along a lower-dimensional manifold $M$ inside $X$, where it continues its existence as a solution to the generalized Seiberg–Witten equations. Proving that this indeed happens would tie together two strands of modern gauge theory in an unexpected way.