Renormalization: Our Greatly Misunderstood Friend

Andrew E. Blechman

September 17, 2002

Abstract

This is a web-paper write-up for a talk I gave for Intermediate Seminar at Johns Hopkins Unversity (172.711-712). I will present a friendly, although slightly technical introduction to the theory of renormalization. Examples from nonrelativistic quantum mechanics and quantum electrodynamics will serve to introduce the concepts. The "Algorithm of Renormalization" will then be presented and explored. Finally, I will present an overview of Wilson's Renormalization Group, and show how we can make the ideas in this paper surprisingly quantitative.

1 Introduction and Example from Nonrelativistic Quantum Mechanics

What is renormalization? It is a buzz word used by many people in all aspects of physics. Popular science authors write about it all the time, and often in a negative light[1]. But what is it really, and why do we need it?

To understand renormalization, we really have to go back and understand "normalization". In quantum mechanics (QM), you have a wavefunction described by Schrodinger's Equation, a second-order partial differential equation. Often, this equation is written in its time-independent form as:

$$\hat{H} \left|\psi\right\rangle = E \left|\psi\right\rangle \tag{1}$$

where $\hat{H}$ is the *Hamiltonian* of the system, E is a constant representing the energy and $\left|\psi\right\rangle$ is the wavefunction. The hat over the Hamiltonian is to emphasize that it is not a number but a linear operator. Once you have found the wavefunction for your system, you can use it to discuss "the probability of measuring something". I am sure many people have heard of Schrodinger's cat, where you are measuring whether the cat is alive or dead, but there are many more practical things you can measure. For example, if your system is a group of particles moving through space with some potential, the wavefunction tells you how probable you are to measure the system with a certain energy.

For this interpretation to be consistent, we must be sure that when you sum over everything, you get unity - there is a 100% chance that something (even nothing!) will happen. That can be expressed by the following equation:

$$\langle\psi|\,\psi\rangle = \int_{-\infty}^{+\infty} |\psi(\mathbf{x})|^2 d^3\mathbf{x} = 1.0 \tag{2}$$

This fixes the constant in front of the wavefunction. Equation (2) is called the **normalization condition**.

Now I want to consider a special method of solving problems in QM called non-degenerate perturbation theory[2]. When solving problems with this technique, you try to express the Hamiltonian of your system in the following way:

$$\hat{H} = \hat{H}^{(0)} + \lambda\hat{V} \tag{3}$$

where $\hat{H}^{(0)}$ is the Hamiltonian of a system that we know exactly how to solve for, and λ is small in some sense. In such cases, we can think of the $\lambda\hat{V}$ term as taking our known system and perturbing it slightly. For that reason, this term is referred to as a "perturbation". Since we know how to solve for $\hat{H}^{(0)}$, I will explicitly write out the (normalized) solutions:

$$\hat{H}^{(0)}\left|n^{(0)}\right\rangle = E_n^{(0)}\left|n^{(0)}\right\rangle \tag{4}$$

$$\left\langle n^{(0)}\middle|\, n^{(0)}\right\rangle = 1.0 \tag{5}$$

where $\left|n^{(0)}\right\rangle$ and $E_n^{(0)}$ are the wavefunction and energy for a state labeled by "n" in the limit $\lambda \to 0$.

Our goal is to find a solution to the entire problem, including the perturbation. In order to do this, we write down the total wavefunction (and the total energy) as a sum of terms, each a different order in λ:

$$|n\rangle = \left|n^{(0)}\right\rangle + \lambda\left|n^{(1)}\right\rangle + \lambda^2\left|n^{(2)}\right\rangle + \cdots \tag{6}$$

In the literature, this series is called a **perturbation expansion**. When doing calculations, we terminate the series at a given order, depending on how accurate we want to be.

You may have noticed a problem at this point. I said that the zero-order states were normalized acording to Equation (5). But the new terms are going to shift this normalization, so that if we want to maintain the probabilistic interpretation of QM, we have to *re-normalize*:

$$|n\rangle_R \equiv Z^{1/2}\,|n\rangle \Rightarrow {}_R\langle n|\,n\,\rangle_R = 1.0 \tag{7}$$

Equation (7) is referred to as the **renormalization condition**, or **RC**.

In nondegenerate perturbation theory, $\left\langle n^{(0)}\,\middle|n^{(1)}\right\rangle = 0$, so if we plug our perturbation expansion (6) in to our renormalization condition (7), we quickly see:

$$Z = 1 + \mathcal{O}(\lambda^2) \tag{8}$$

In other words, renormalization enters the picture at *second* order in the perturbation expansion. This result turns out to be true in many instances, as we will see in the next example of Quantum Electrodynamics.

As a final point, realize that I have said nothing about whether or not a perturbation expansion even makes sense mathematically. You are expanding the wavefunction and energy in powers of a constant, and mathematicians might find this very disturbing. It turns out that you are right to be disturbed, and it has been shown that a perturbation expansion does not converge in general. However, it can be shown that in most cases it succeeds as an "asymptotic expansion", and hense makes sense as long as you don't try to go out too far[3]. We will not worry about this point in this paper.

2 Example from Quantum Electrodynamics

Quantum Field Theory (QFT) is where renormalization becomes truly interesting, and a little suspect! At a first glance, nearly all forms of field theory are sick. The reason is that when you attempt to perform a perturbation expansion, you find that beyond the leading order you calculate infinite quantities. Certainly it does not make sense to say that the probability for a process to occur is not only greater than 1.0, but *infinite*; perhaps we should just pack up and go home.

No one wins a Nobel prize this way! Instead, people began to realize that the reason QFT results blow up is because you need to renormalize. In this section I will quote a very straightforward example. Then in the rest of the talk, I will discuss how we go about renormalizing our theory.

Consider one of the simplest processes in quantum electrodynamics (QED), the quantum theory of radiation: an electron annihilating a positron (or antielectron) to create a muon-antimuon pair. We write the process as $e^-e^+ \rightarrow \mu^-\mu^+$. There is a beautiful technique developed by Richard Feynman for writing down processes of this sort. Represent the particles by solid lines, and photons (the quanta of the electromagnetic field) as a swiggly line. The only allowed vertex in QED is two particles connecting to a photon, so the Feynman

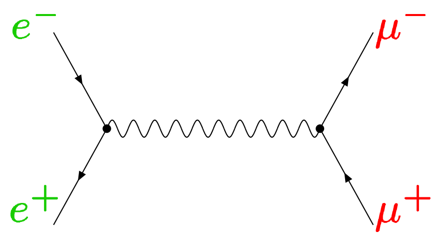

Figure 1: $e^-e^+ \to \mu^-\mu^+$ at leading order

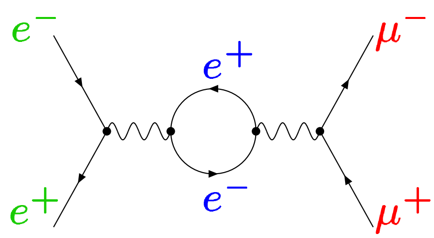

Figure 2: $e^-e^+ \to \mu^-\mu^+$ with one loop

diagram for this process is given by Figure (1), where time flows left to right and antiparticles get backwards arrows by convention. Because of a lack of symmetry, this is the only way this process can occur, so we need only write down this diagram. Each diagram represents a complicated mathematical expression, called the **amplitude** for the process. The order of these expressions can be inferred by the number of vertices in the diagram. In this case, there are two vertices, and so this process is of order e^2, where e is the charge of the particles. Equivalently, we can say that this amplitude is order α, where $\alpha = \frac{e^2}{4\pi}$ is the fine structure constant in natural units ($\hbar = c = 1$).

Now consider the next term in the perturbation expansion. In terms of Feynman diagrams, these terms can be thought of as diagrams with loops, so that they have more vertices but do not change the external states. An example of such a diagram is in Figure (2), but realize that this is not the only diagram. When you create a loop, you are creating a particle-antiparticle pair. This pair has energy and momentum, and in the spirit of the superposition principle of QM, you must integrate over all these momenta. This is where the problem lies.

Each loop involves one integral over four-momenta, and each particle contributes a factor of inverse four-momentum to the integrand. Hense a diagram of the form in Figure (2) has an integral which goes as momentum squared, and this diverges as you integrate to infinity.

This kind of analysis is called "power counting". It turns out not to be an acurate way to decide whether an integral converges or not. For example, the above suggests that this is a quadratic divergence, but if you actually did out the integral you would find that it diverges logarithmically. However, power counting does give you a first guess behavior of your integral, and in this case, it is true that the integral diverges.

Just because the integral diverges does not mean that we should give up! Rather, it means that we must renormalize to get a finite result.

3 Algorithm of Renormalization

We have seen what renormalization is in the context of perturbation theory, and we have seen how without renormalization, even the simplest processes in QED would blow up. Now we must discuss renormalization itself - how is it accomplished, and how should it be interpreted. It is the first of these questions I will try to answer in this section. In the next section, I will discuss the interpretation.

The Algorithm of Renormalization proceeds in two steps. The first step is to perform the integration and separate out the divergent pieces, and the second step is to get rid of these divergences. The first step is often called **regularization**, since we are "regulating" the divergence in the integral. This is opposed to the second step of removing the divergence, which I will call **renormalization**. I would like to discuss each of these steps separately.

3.1 Regularization

The goal of regularization is to explicitly calculate the divergent integral:

$$I \equiv \int_0^\infty d^4k F(k) \tag{9}$$

This can be done a number of ways. Furthermore, the end result should not depend on which regularization scheme you chose. The idea will always be to reparametrize the integral in terms of a parameter which I will call a **regulator**. After I have expressed the integral in this way, I will take the physical limit where the result returns to the original integral. Terms that vanish, I will ignore. Terms that blow up, I will need to get rid of. The end result will be something finite.

In this talk I will present four different regularization schemes. There are many more, but these are the most commonly used. Notice that every term in this section can also depend on physical quantities such as mass, charge and external momenta. I am supressing this dependence, as it is not relevant to the regularization scheme.

- **Momentum Cutoff.** We are evaluating integrals that have the form of Equation (9). This integral diverges at the large limit. Then perhaps the most obvious choice for a regularization scheme is not to integrate to infinity, but to a very large momentum, parametrized by the greek letter Λ:

$$I \to I_\Lambda \equiv \int_0^\Lambda d^4k F(k) \tag{10}$$

  I_Λ is certainly convergent, and becomes I in the limit $\Lambda \to \infty$. We can do this integral to get the general result:

$$I_\Lambda = A(\Lambda) + B + C\left(\frac{1}{\Lambda}\right) \tag{11}$$

  where in the physical limit $\Lambda \to \infty$, A is divergent (either power-law or logarithmically), C vanishes, and B is independent of Λ and hense remains finite. We can immediately drop C, and we are left with a piece that diverges and a piece that is finite. If we can only figure out a way to get rid of the divergent piece, we can take the limit and get a finite answer. This is exactly what we will do in step 2. But before getting there, I'd like to present a few more regularization schemes.

  Although momentum cutoff regularization is probably the most obvious choice for a regulator, it is rarely the best one. The reason is that the momentum cutoff dependence almost always violates an important symmetry of the theory such as gauge invariance, which is needed to make sure quantities cancel correctly. Therefore, unless we want a qualitative understanding of the diagram, we almost never use momentum cutoff regularization in practice.

- **Dimensional Regularization**. DimReg, as oposed to momentum cutoff, is one of the most useful and least intuitive regulators. The reason for its usefulness is that it preserves gauge invariance and keeps all the symmetries of the theory manifest.

  In DimReg, we replace our integrals with:

$$I \to I_d \equiv \int_0^\infty d^dk F(k) \tag{12}$$

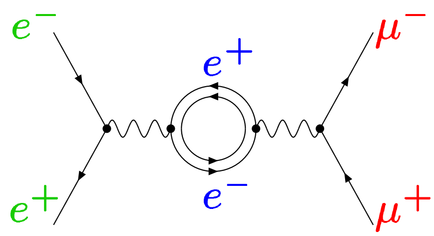

Figure 3: Pauli-Villars counterdiagram

where d is the dimension of our measure. This is *not* to be confused with extra dimensions and string theory - the physical result corresponds to $d = 4$. To that end we can perform the integral in d spacetime dimensions, and replace $d \to 4 - \epsilon$, so that the physical limit is $\epsilon \to 0$. We are left with:

$$I_\epsilon = A(\epsilon) + B + C\left(\frac{1}{\epsilon}\right) \tag{13}$$

This time, however, it is A which we can drop immediately in the physical limit. Again, the goal of step 2 will be to find a way to remove the divergent piece, leaving the finite piece over when we take the limit $\epsilon \to 0$.

- **Pauli-Villars Regularization**. PV regularization involves the intriguing philosophical assumption that there is more physics going on than we see. When writing down the amplitude, assume that there is another diagram with a loop, this time with a particle of mass M larger than anything else in the theory; see Figure (3). However, this diagram enters *with the wrong sign!* Hense you are subtracting a diagram from your theory. Diagrams with massive loops behave as $\frac{1}{M^2}$, so in the limit $M \to \infty$ this diagram does not contribute. However, while keeping M finite it will help to cancel divergences. The logic works exactly like momentum cutoff regularization, where we must remove terms that diverge with M and drop terms that vanish in the physical limit, leaving a finite result.

- **Lattice Regularization**. Lattice regularization is a very different beast than the regulators we have been talking about, but I feel obligated to talk about it here because it is truly very beautiful. In lattice regularization, you assume that the universe is not a continuum, but rather a discrete lattice. Now all of the integrals are actually sums, and you never integrate to infinity since there is a natural cutoff, namely the lattice spacing a. After performing the finite sums, you take your theory off the lattice by going to the physical limit $a \to 0$. Again, you get terms that diverge with a that you must remove by the renormalization step.

Lattice regularization is very different from any of the other schemes because it is not perturbative. In other words, you use lattice regularization when you are trying to solve a problem *without* using perturbation theory. This is very useful in theories with the strong force described by quantum chromodynamics (QCD), since the perturbative regime of QCD is very limited. Lattice QCD has been very successful in predicting many results of low-energy QCD, including confinement, hadron masses and form factors[4].

3.2 Renormalization

The technique of regularization gave us a way to "parametrize the infinities". Now we must develop a way to get rid of these infinities. This is the step of renormalization. For the sake of notation, I will call my regulator Λ; this does not mean that I have used momentum cutoff regularization.

Up to this point, the integral we have been considering can be denoted the following way:

$$I \equiv I(m, \alpha, \Lambda) \tag{14}$$

This quantity blows up in the physical limit. However, it turns out that we can capture this divergence if we make a clever shift of the physical parameters:

$$\begin{aligned} m &\to m(\Lambda) \equiv m + \delta m(\Lambda) && (15) \\ \alpha &\to \alpha(\Lambda) \equiv \alpha + \delta\alpha(\Lambda) && (16) \\ I(m, \alpha, \Lambda) &\to I(m(\Lambda), \alpha(\Lambda)) && (17) \end{aligned}$$

There is nothing different about this result, except that I have absorbed all of the divergent behavior into the physical parameters, so that I is no longer *explicitly* divergent, but merely dependent on divergent but physical quantities.

Now I hear you say: "But now we're worse off than before! After all, we know that the electron is not infinitely massive or has infinite charge!" But what you are forgetting is that I have not yet specified my renormalization conditions yet. In fact, if I chose for my RC:

$$m(\Lambda) \to m_R \qquad\qquad \alpha(\Lambda) \to \alpha_R \tag{18}$$

in the physical limit (the "R" stands for "renormalized"), then our final result is simply $I(m_R, \alpha_R)$ - a finite answer! m_R and α_R are now just the quantities we measure for the electron mass and charge respectively. We have litteraly "swept the infinities under the rug" to extract the UV-finite solution.

Before you label this as ridiculous, realize that QED has used renormalization all the time, and its results have been tested to as many as fourteen decimal places. That is the best known confirmation of *any* theory of physics. Surely we must be doing something right!

As a final point, I wanted to mention that our final (finite) results cannot depend on the regulator. However, you may be concerned about the process of absorbing the regulator into the physical quantities. For example, do you absorb any of the finite part of your regulated integral? Each regularization scheme gives you a different finite part, so how are you to know what to drop and what to keep? The answer is that you must explicity state these details when you give your final answer. This final step is often called the "subtraction scheme". So when you quote a final renormalized answer, you must state what subtraction scheme you used to renormalize the obervable quantities. Some of the more common subtraction schemes are pole subtraction (P), where you extract the mass directly from the amplitude of a particle with zero momentum; minimal subtraction (MS), where you subtract *only* the divergent part of the amplitude; and modified minimal subtraction ($\overline{MS}$), where you subtract certain additional finite terms from the MS scheme. There are many others. In practice, you must be careful to be consistent with your subtraction scheme in your calculations.

4 Theory of Renormalization

You might still be bothered by many things in the previous section. When we renormalize physical quantities such as charge and mass, you might be thinking that these quantities are observable and are not infinte. So how can you get away with making them divergent and then ignoring it?! The answer to that question is actually deeper than it first seems.

First of all, there is a flaw to the skeptic's argument that the electron is not infinitely massive or carries infinite charge. In fact, according to QFT, it does! The reason we don't see it is subtle but beautiful. If the electron has infinite charge, then it has an infinite amount of energy from the electromagnetic field. This energy manifests itself by the uncertainty principle which says that the field is allowed to create and destroy particles in very short times; such particles are called "virtual particles". With this huge amount of energy, the field is able to produce many particles with charge all around the electron. But because these virtual particles are charged, they line up with the field and dampen the strength, analogously to dielectrics in classical electrodynamics. Hense as you go further away from the electron, its effective charge becomes weaker due to this dielectric effect, thus lowering the charge of the electron to the values we measure.

Ah, but in that case, shouldn't the electron's charge get larger and larger as we get closer and closer to it, cutting through this quantum dielectric? The answer is yes, and perhaps even more amazingly, this is precisely what happens! In the everyday world, we measure $\alpha = \frac{1}{137}$, but at high-energy accelerators such as the Tevatron at Fermilab, we measure $\alpha = \frac{1}{128}$ - this is a real effect[5].

In the past, this effect has been calculated directly by deriving the "Uehling Potential" which is the quantum correction to the Coulomb potential. However, in the past thirty years or so, physicists have developed a much more powerful technique for describing these results in a beautifully elegant and intuitive way. This technique was pioneered by physicists K. Wilson, M. Fisher, L. Kadanoff and others. The technique is called the **Renormalization group**.

The renormalization group is a very complicated object, but I will just say a few things about it. In general, the idea is to do everything that we have been doing, only now our RC will generally depend on the scale of our experiment; call it μ. This scale is referred to as a **subtraction point**; its value depends on the scale of the experiment as well as the subtraction scheme. Then we can ask: "How do our physical parameters depend on our subtraction point?" We can write down a set of differential equations, called the "renormalization group equations" that try to answer how our couplings and masses evolve with changing subtraction point. This gives us the results I said above.

It is this technique that has led to the discovery of **asymtotic freedom**, the key quality of QCD, where the forces get *weaker* as the subtraction point increases. This allows for a perturbative analysis of QCD at high energies, when perturbation theory fails at low energies[4]. In addition, the renormalization group has helped to solve a number of questions in statistical mechanics, such as the behavior of magnets, liquid crystals and general phase transitions, just to name a few[6].

5 Discussion

"Renormalization" is a word that has been given the evil eye by mathematicians, philosophers and popular science writers ever since it was first used to regulate infinities prevalent in quantum field theories. For a long time, many physicists have also looked down on it as a necessary although unattractive proceedure. However, as quantum field theory becomes time tested again and again, it becomes harder to simply write off renormalization as a bad idea.

When Wilson, et al. published their derivation of the renormalization group, physics underwent a spectacular shift in philosophy. No longer was renormalization a necessary evil, but a *requirement!* It provided an entirely new way of interpreting ultraviolet divergences. The general philosophy of quantum field theorists is now that any given theory of physics has some energy scale where the theory breaks down. Renormalization not only allows you to perform calculations below that scale, but through the renormalization group equations, tells us where that scale is! This allows people to predict where to find new physics. For example, this is how people predicted the top quark mass.

Renormalization has not only helped us to explore the perturbative regime of quantum field theories, but has also given us great insight into the nature of how physical theories must scale with energy. It has given us a way to deeply probe the nature of the theory itself. While mathematicians and philosophers continue to call it a problem, physicists have learned that it truly is our deeply misunderstood friend.

References

[1] There are many books out there, but one of my favorite is Gribbin, J., *In Seach of Schrödinger's Cat*, Bantam Doubleday Dell Pub., 1985.

[2] Sakurai, J. *Modern Quantum Mechanics* (Revised Ed), Addison-Wesley Publishing Co., Inc., 1994.

[3] Dyson, F., *Phys. Rev.* **85**, 631, 1952.

[4] Blechman, A., *Coloring the Universe with Quantum Chromodynamics*, http://www.pha.jhu.edu/~blechman/papers/qcd.ps

[5] For details on all of the QFT results, see Peskin, M. and Schroeder, D., *An Introduction to Quantum Field Theory*, Perseus Books Pub., 1995.

[6] For a review of the renormalization group in statistical mechanics, see Chaikin, P. and Lubensky, T., *Principles of Condensed Matter Physics*, Cambridge University Press, 1995.