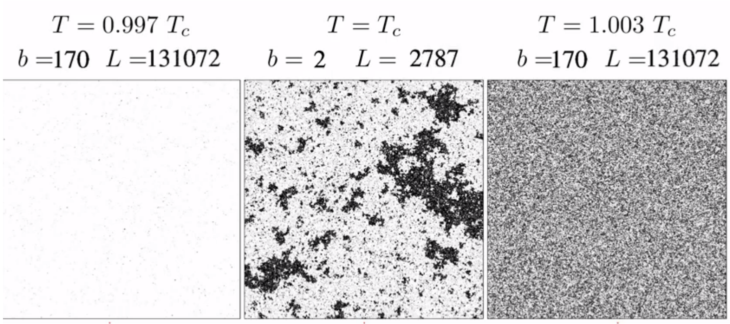
The Renormalization Group

Marcus Berg, March 6, 2017

1 Basic history: particle physics and condensed matter physics

The Wikipedia page on the renormalization group goes back to Pythagoras, and there was certainly some early related work by particle physicists on renormalization in the 1950s. But the key steps to the modern renormalization group are: Kadanoff in 1966 [3], Callan and Symanzik in 1970 [4, 5], and then in 1971 Kenneth Wilson[1] began a synthesis of the earlier work. Wilson made the renormalization group into a more universal program that led to his 1974 solution of a ten-year old problem in cond-mat (condensed-matter, what used to be called "solid state physics"), called the "Kondo problem", as he summarizes in his Nobel lecture [6].[2] Wilson wrote a beautiful popular-science article about the renormalization group in [8]. Of course nowadays there is a one-minute YouTube video about it [9], where I got this picture:



So, first condensed matter physics for orientation. In the free book Gould & Tobochnik [1], there is first a discussion of the simple process of *percolation*, then (in Ch. 9.5) the renormalization group for the 1D Ising Model of interacting magnetic spins in a one-dimensional chain. In my statistical mechanics notes I discuss the precursor of this, the mean field theory approximation, where the magnetic effect of all other spins on a single spin is collected into a modification of the background magnetic field. In Problem 9.25, Gould & Tobochnik discuss the Kosterlitz-Thouless phase transition (Nobel prize 2016). Although they don't discuss it there, Kosterlitz applied renormalization group method to their transition in 1974 [2], which was part of their cond-mat Nobel prize.

2 Renormalization group in quantum field theory

2.1 Summary

One way to state Wilson's insight is that amplitudes are not organized by graphs, but by scale. This is the front page of Peskin & Schroeder. It basically means that the coupling constant g counts the number of loops, but the energy scale Q (which can be thought of as being set by experiment[3]) affects how

[1] Usually his first name is emphasized the first time you bring him up, since there are several other Wilsons in physics.

[2] *"The result was a recursion formula in the form of a nonlinear integral transformation on a function of one variable, which I was able to solve by iterating the transformation on a computer."* I find it pretty amazing that any single researcher did anything on a computer in 1970. Paul Ginsparg in his reminisces about creating the arXiv [7] writes "My thesis advisor Ken Wilson repeatedly promoted to us the need for massive parallel processing". So in an admittedly far-fetched sense, the arXiv is a spin-off of the renormalization group,

[3] Why is it called Q and not for example E? This is because I have the proton collider at CERN in mind, where Q is the scale of momentum transfer between quarks and gluons in the colliding protons, which is not automatically set by the energy of the injected protons. The actual momentum transfer can only be determined with hindsight, when experimenters have picked out the "interesting" proton-proton collisions. But given this hindsight, Q can indirectly be thought of as the energy scale of the experiment.

many loops you should keep for given accuracy. A practical discussion of this for experimentalists is given by the Particle Data Group — more about this below.

2.2 Discussion

A simple important point that in my experience is sometimes not stated clearly is that if we could compute something *exactly*, we could worry less about "given accuracy" in the little summary above. Let me try to explain this.

There are lots of explicit equations about all these things in the textbooks. But it's easy to get lost in the details. So here's a simplistic attempt to illustrate the point. The difference between the exact and truncated cross sections as functions of coupling g and energy Q can be illustrated as[4]

$$\sigma_{\mathrm{exact}}(g,Q) = \underbrace{|\text{———————————}}_{\text{perturbative}}\underbrace{|\text{———}|}_{\text{nonperturbative}} \tag{2.1}$$

$$\sigma_{\mathrm{trunc}}(g,Q/M) = \underbrace{|\text{———————}|^{M}---}_{\text{perturbative}}\underbrace{|---|}_{\text{nonperturbative}} \tag{2.2}$$

First, I have emphasized that in addition to an all-loop-order perturbative result (sometimes itself incorrectly called "exact"), there can be an additional piece that is nonanalytic in g, like $e^{-1/g}$, which is nonperturbative[5]. The exact result σ_{exact} does not necessarily depend on the arbitrary renormalization scale M and can be matched to experiment. In condensed matter physics this is conceivable to do directly.

In particle physics, roughly speaking the only models where σ_{exact} is known are the ones that are not realistic enough to match to experiment, e.g. in low dimension (e.g. Gross-Neveu model), or with strong symmetry assumptions (like maximally supersymmetric Yang-Mills theory, where partial results are known). By contrast, a realistic theory is QCD (quantum chromodynamics), the theory of the strong nuclear interaction. A typical strong-interaction QCD cross section measured at the LHC at CERN is σ for ZZ production, which is 7 pb (picobarn, see my particle physics notes).

To compute such cross sections from theoretical considerations, we consider σ_{trunc} at some order in perturbation theory, and let us also ignore nonperturbative effects. The symbol M in σ_{trunc} above is meant to represent that we truncate at a given loop level L (where the perturbative contribution is of order g^L) and introduce a scale M at which we define the cross section σ. Typically a loop correction is $g\log(Q/M)$, and at some experimentally accessible scale $Q = M$ we have $\log(Q/M) = \log 1 = 0$, so in practice the tree-level value (i.e. for $g \to 0$) is the one that it matched to experiment at that energy scale.

The first time one hears about the arbitrary introduction of the unphysical reference scale M, many students find it almost unbearable. (I know I did.) It is related to another even more unbearable fact, that finite parts of amplitudes depend on an arbitrary choice of what to cancel along with any possible infinities ("renormalization scheme").[6] As you might guess from the way I stated this, the

---

[4] This can be confusing as the textbooks often give a (useful) dimensional analysis argument that dimensionless observables cannot depend on the energy Q by itself but can only depend on the combination Q/M, where M is an unphysical reference scale that is inserted by hand. This useful argument has several caveats: at high energy there might be a new physical fundamental energy scale (in string theory, the string scale M_s, but it could equally well be the GUT scale M_{GUT} or the Planck scale M_P, or just the scale of heavy quarks M_Q). Or, the observable of interest may simply not be dimensionless: σ has dimension of area so the naive dependence is as Q^{-2} without the unphysical M.

[5] Assuming something about the value of the coupling , it is easy to make a naive estimate that nonperturbative effects $e^{-1/g} \sim g^L$ for some number of loops L, e.g. for $g \sim 0.1$ they become comparable to perturbative contributions at $L \sim 5$ loops. Examples of nonperturbative contributions in quantum field theory include *instantons* and *renormalons*. The latter are potentially more harmful (technically, they obstruct Borel resummability of the pertubation series), but may be avoided under some circumstances. Both are well described in Weinberg's book.

[6] In fact the arbitrariness is even worse than I gave the impression of above: loop corrections to amplitudes can vary in sign, so the cross section could either shrink or grow when we compute more loop corrections. It's OK to ignore this issue for now and imagine that perturbation theory converges nicely, and we accumulate positive corrections to ultimately reach some nice finite "end" value for σ like 7 pb. To be honest, this is not the case in any really physical example, but the issue of infinities is logically distinct from conceptual understanding of the renormalization group, see below.

two issues in a certain sense "cancel" each other. In fact this is how Peskin & Schroeder leads into the discussion of the renormalization group (around eq. (11.80)), but I find that explanation somewhat confusing.

I included some comments about QCD vs. experiments above to help make this more concrete: in particle physics there is an enormous amount of data that is more or less understood, and the relevant Nobel Prizes have been awarded to theorists like Gross et al (2004) for theoretical calculations that are in fact matched to experiments, like the strong coupling g_s, or the combination $\alpha_s = g_s^2/(4\pi)$. This is relatively concisely summarized in the Review of Particle Physics by the *Particle Data Group* (pdg.lbl.gov), where all key results in particle physics are available online[7]. Their review of QCD [10] does not attempt to teach QCD, but discusses the scheme dependence of higher-order terms in the strong coupling α_s for practically-minded experimentalists. Let me come back to this below after I introduce the basic equation they solve.

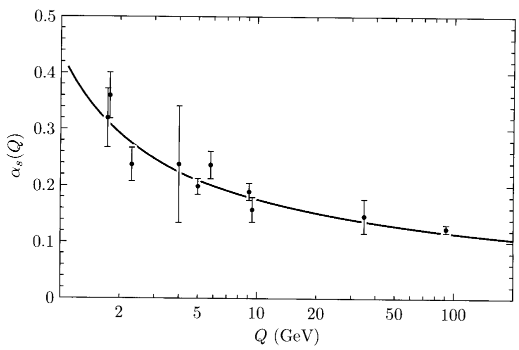
2.3 The renormalization group equation

We will use the version of the renormalization group equation called the *Callan-Symanzik equation*:

$$\left(M\frac{\partial}{\partial M} + \beta(g)\frac{\partial}{\partial g} + n\gamma\right) G^{(n)}(x_1, x_2, \ldots, x_n; g, M) = 0 \ . \tag{2.3}$$

that tells us how an n-point correlation function $G^{(n)}(x_1, x_2, \ldots, x_n; g, M)$ depends on renormalization reference scale M. The functions β and γ can be calculated in perturbation theory and are specific to each theory.

Essentially, the equation instructs us that the object of interest does not depend on g and M independently, but that a change in one of them can be compensated by a change in the other, so that we are effectively following a curve in the (g, M) plane instead of exploring the entire plane. But really, why is it better to use this equation than to do the naive thing: compute at fixed g and fixed loop order L, and vary the energy of the experiment Q as you wish? The answer is the Peskin & Schroeder cover: only increase accuracy (loop order) when we need it, and then use information about lower loop orders as boundary condition of this differential equation to infer something about some dominant feature (usually *leading logarithms*) of the next loop order. The result for $\alpha_s(Q)$ is Peskin & Schroeder Ch. 17.6 (there are more recent plots from the PDG, but I find this one the clearest):

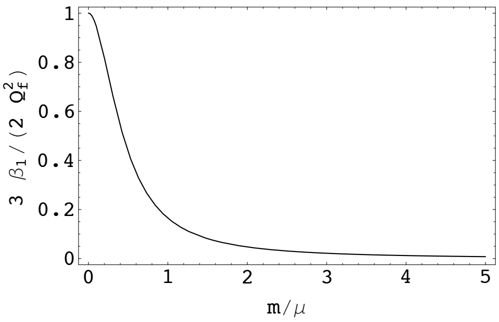


An analogy I personally like is "adaptive mesh refinement" in computing, that you have already used every time you solved an ODE numerically in Maple or Mathematica: where a function changes quickly in some region, the algorithm adds more mesh points in that region only, i.e. increases computational effort. Changing quickly means having structure on relatively short distance, which corresponds to high frequency f by Fourier transform, or high energy by $E = hf$. So if you want to probe high energy, only increase effort (compute to higher loop order) for those observables that are sensitive to an increase in energy ("logarithms become large"). This clever strategy is called the "renormalization group improvement" of perturbative quantum field theory. An explicit example is in Weinberg Vol.2, section 18.8.

[7] I casually note the fact that I'm credited in the Review of Particle Physics as a "consultant" [11] for the Particle Data Group (PDG) up until 2001. Ask me why!

3 Decoupling

In fact due to the above issues of truncation, despite the nice plot it is fair to say that the strong coupling constant α_s in experiments is itself not an "observable" in any real sense. Related to this is that the beta function $\beta(g)$ itself is in general not physical, and may in fact even change discontinuously as we cross an energy threshold. (That is why I insisted on σ above: cross sections are observable[8].) As Pich [14] argues, this can be thought of as being due to the standard practice of computing β in the technically convenient but somewhat unphysical scheme of dimensional regularization with minimal subtraction, or "$\overline{\text{MS}}$". In a "more physical" scheme like one that Pich introduces, the QED beta function for mass m looks like (Pich defines $\beta = \beta_1 \cdot \alpha/\pi$, so that $3\beta_1/(2Q_f^2) = 1$ in the massless theory with Q_f fermions):



Here things are more like one might expect: the effect of heavy fermions on the renormalization group evolution of low-energy parameters (the β function) decreases when the fermion is made very heavy. In other words, particles "decouple" when they become much too heavy to produce at any realistic collider experiment, as one might have hoped. (Of course, strictly speaking one arbitrary choice cannot really be "more physical" than another arbitrary choice in any general sense, but it can capture an issue like this more clearly.)

In fact, an oft-quoted "physics theorem" about decoupling is by Appelquist and Carazzone already in 1975: *"heavy fields decouple at low momenta except for their contribution to renormalization effects"* [16]. A cautionary tale about this is the Gastmans-Wu-Wu paper from 2011 about the Higgs decay to two photons [17] where they claimed the old Standard Model calculations don't satisfy the Appelquist-Carazzone decoupling theorem. The Russian response [18] explains very clearly why Gastmans et al (and Gastmans is a well-known expert in an illustrious group of Dutch physicists!) misunderstood the theorem, *"as if the passage of time negates the knowledge of the past"*.

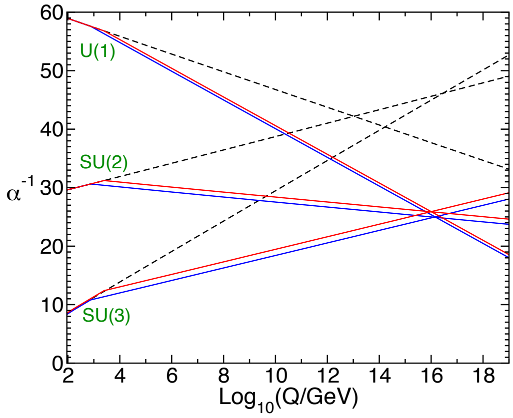
4 Matching

However, now there is a confusing point: if we use a mass-independent subtraction like the standard $\overline{\text{MS}}$, how *do* we cross a mass threshold where new physical states become relevant, like the mass of a heavy quark? You use a different β function above and below the threshold, this is called "matching". For example, the number of experimentally accessible quarks depends on energy (three for low energy, six for high), and the beta function depends on the number of quarks. I expect that Schwartz describes this well, but it is not very clear in Peskin & Schroeder. I like Georgi's explanation [12], or why not my explanation in the appendix of a dark matter paper [13].

[8] As always, even this statement requires qualification: as Polchinski warns in his book, we should specify whether we mean"inclusive" versus "exclusive" cross sections depending on what processes you count (this is discussed in section 9.2 in the QCD review from PDG), and infrared (low-energy) divergences due to massless particles like gluons make some cross sections infinite. The latter issue goes all the way back to the Rutherford cross section for scattering alpha nuclei from gold atoms, that is infinite at zero scattering angle, but for a sort-of-physical reason: the Coulomb force has infinite reach, related to the fact that the photon is massless. But there are ways to only consider infrared-safe observables, as discussed in Weinberg and more recently by others [19]. Arkani-Hamed's talk at a recent Simons Center workshop on string theory and scattering amplitudes I went to, available online, discusses this in great detail.

5 Grand unification

Some students have already seen the gauge coupling unification diagram (fig.6.8 in Martin [20]):



Note that the horizontal axis is a logarithmic scale. The same figure is fig. 22.1 in P&S, but there it is somewhat harder to see the kinks (discontinuous changes of slope) in the couplings around where the gauge groups are shown in green above. The kinks are precisely what I was talking about above: that is where there is matching between the theory below the scale $M_{\rm susy}$ of superpartner masses and the theory above that scale – and the change in slope is pretty big!

Unification of gauge groups as in the dashed (nonsupersymmetric) lines was first studied by Georgi and Glashow in 1974, as discussed in Peskin & Schroeder. (They describe the Georgi-Glashow paper as "remarkable hubris" in section 22.2). The basic idea is that the Standard Model product group $SU(3) \times SU(2) \times U(1)$ is unified to the bigger group $SU(5)$. (Gravity is still left out.) In later versions the unification group became $SO(10)$ or even E_6 or E_8, as discussed in e.g. Polchinski Ch. 11. The renormalization group evolution of coupling constants so they can meet is a basic ingredient in all of this. A related important point is that the evolution drawn above can receive finite but important corrections close to the "threshold" of creating physical new particles, the scale where coupling cross. Such additional finite corrections are called "threshold corrections". The equation that governs them is called the Georgi-Quinn-Weinberg equation and is discussed in Polchinski Ch. 18.

6 Confinement

One of Wilson's expressed goals was to understand the confinement problem for quarks [15] (his paper has over 4000 citations which is a lot for an old paper!). The problem remains unsolved, but it is true that it was made clearer by the calculation of $\beta(g)$ for QCD. The point is: QCD has $\beta(g_s) < 0$ (Nobel Prize 2004), which means the coupling $g_s \to 0$ at high energy $Q \to \infty$ ("asymptotic freedom"), as you saw in the picture above. But then conversely, $g_s \to \infty$ for some *low* energy scale, which turns out to be $Q \to \Lambda_{QCD} \sim 250$ MeV, the scale of light hadrons. This signals confinement, though if $g_s \to 0$ we can no longer compute in perturbation theory. The best current hope seems to be to understand confinement in simplified (supersymmetric) theories through duality, the basic example being Montonen-Olive duality (Polchinski Ch. 14).

6.1 Does QED exist?

As Nahm said in a talk, "perhaps the reason that QED is not so interesting mathematically[9] is that it doesn't exist". For $\beta(g) > 0$ as in QED, there is a well-known problem known as a Landau pole, which is the opposite of the problem for QCD at low energy. For some extremely high but finite energy (over 10^{200} GeV!), the QED coupling (electric charge) may diverge due to renormalization group

[9] What he really said was "does not have an interesting moduli space"

evolution. This is not the same as infinities of perturbation theory, this is a more or less "physical" statement, which instead can be interpreted as saying that QED by itself doesn't make sense. However also this statement is not clear, for example the issues are affected by duality, as for QCD above. More physically, if QED is unified to an asymptotically free ($\beta < 0$) theory as in the plot above long before the scale of the Landau pole, this issue never becomes relevant.

7 Is renormalization related to infinities?

No. Weinberg tries to be particular clear about this. Even in a theory where all momentum integrals are finite, there is still renormalization. As Polchinski writes, the "true meaning" of the renormalization group is to say how the theory varies with scale, not to solve problems with infinities. I personally agree with this, but I'm not sure it's widely agreed upon or appreciated in high-energy theory outside of Weinberg's hallway in Austin, Texas, where Polchinski used to have his office.

In fact, a somewhat more "modern" version of the Callan-Symanzik equation is the Wilson-Polchinski equation from 1984 [23]. In 2009, Polchinski described it like this [22]: "The statement of renormalization becomes just this differential statement: How many positive and how many negative eigenvalues does the flow operator have? Since the flow is differential, it only has a narrow range of energies. Everything is finite. There's no UV divergences, there are no IR divergences. And so you can say the eigenvalues at small coupling have to be very close to the values at zero coupling because there's no place for any big effects to come from. And that's the whole content."

In another context, Polchinski described his early work like this: *"...my own misspent youth. I used to focus too much on rigor and formalism, and have become a much more creative and productive scientist since learning, very slowly, to see through these to the physics."*

8 A little math

I can't help point out that the renormalization group β function of the nonlinear sigma model (discussed at length in Peskin & Schroeder) was studied by for example Friedan in 1980 [21] related to some work in string theory, and this in turn lead to what is known as "Ricci flow" in mathematics:

$$\partial_t g_{ij} = -2R_{ij} \tag{8.1}$$

where t is "renormalization group time" (essentially $t = \log(Q/M)$ from above), g_{ij} is a metric and R_{ij} is a Ricci tensor. The Ricci flow equation looks a little bit like a heat equation. Heat diffuses so that uneven temperature distributions are smoothed out, but here it is *curvature* that is smoothed out. This lead to Perelman's solution of the Poincaré conjecture, the only one of the seven Clay problems to be solved so far.

References

[1] Gould, Tobochnik, "Thermal and Statistical Physics", `http://stp.clarku.edu/notes/` Ch. 9, "Critical Phenomena and the Renormalization Group".

[2] J. M. Kosterlitz, "The critical properties of the two-dimensional xy model" (1974), J. Phys. C: Solid State Phys. 7, 1046

[3] L. P. Kadanoff, "Scaling laws for Ising models near T_c," Physics **2** (1966) 263.

[4] C. G. Callan, Jr., "Broken scale invariance in scalar field theory," Phys. Rev. D **2** (1970) 1541. doi:10.1103/PhysRevD.2.1541

[5] K. Symanzik, "Small distance behavior in field theory and power counting," Commun. Math. Phys. **18** (1970) 227. doi:10.1007/BF01649434

[6] K. Wilson, Nobel Lecture (1982), `https://www.nobelprize.org/nobel_prizes/physics/laureates/1982/wilson-lecture.html`

[7] P. Ginsparg, "It was twenty years ago today ...," arXiv:1108.2700 [cs.DL].

[8] K. G. Wilson, “Problems In Physics With Many Scales Of Length,” Sci. Am. **241** (1979) 140. doi:10.1038/scientificamerican0879-158

[9] D. Ashton, “The Renormalization Group”, YouTube video, `https://www.youtube.com/watch?v=MxRddFrEnPc`.

[10] S. Bethke, G. Dissertori, G.P. Salam (for the Particle Data Group), “Quantum Chromodynamics, Review” (2016) http://pdg.lbl.gov/2016/reviews/rpp2016-rev-qcd.pdf

[11] Review of Particle Physics, “Introduction” (2001) http://pdg.lbl.gov/2001/textrpp.pdf

[12] H. Georgi, “Effective field theory” http://www.people.fas.harvard.edu/ hgeorgi/review.pdf

[13] M. Berg, J. Edsjo, P. Gondolo, E. Lundstrom and S. Sjors, “Neutralino Dark Matter in BMSSM Effective Theory,” JCAP **0908** (2009) 035 doi:10.1088/1475-7516/2009/08/035 [arXiv:0906.0583 [hep-ph]].

[14] A. Pich, “Effective field theory: Course,” hep-ph/9806303.

[15] K. G. Wilson, “Confinement of Quarks,” Phys. Rev. D **10** (1974) 2445. doi:10.1103/PhysRevD.10.2445

[16] T. Appelquist and J. Carazzone, Phys. Rev. D **11** (1975) 2856. doi:10.1103/PhysRevD.11.2856

[17] R. Gastmans, S. L. Wu and T. T. Wu, “Higgs Decay $H \to \gamma\gamma$ through a W Loop: Difficulty with Dimensional Regularization,” arXiv:1108.5322 [hep-ph].

[18] M. Shifman, A. Vainshtein, M. B. Voloshin and V. Zakharov, “Higgs Decay into Two Photons through the W-boson Loop: No Decoupling in the m_W -> 0 Limit,” Phys. Rev. D **85** (2012) 013015 [arXiv:1109.1785 [hep-ph]].

[19] N. Arkani-Hamed and J. Maldacena, “Cosmological Collider Physics,” arXiv:1503.08043 [hep-th].

[20] S. P. Martin, “A Supersymmetry primer,” Adv. Ser. Direct. High Energy Phys. **21** (2010) 1 [hep-ph/9709356].

[21] D. Friedan, “Nonlinear Models in Two Epsilon Dimensions,” Phys. Rev. Lett. **45** (1980) 1057. http://www.physics.rutgers.edu/∼friedan/papers/PRL_45_1980_1057.pdf

[22] AIP History, Interview with J. Polchinski (2009) https://www.aip.org/history-programs/niels-bohr-library/oral-histories/33729

[23] J. Polchinski, “Renormalization and Effective Lagrangians,” Nucl. Phys. B **231** (1984) 269. doi:10.1016/0550-3213(84)90287-6