LGT, Unit 4, Lectures 1 and 2: RG and Scale Setting

Jake Sitison

October 15, 2024

Learning Goals

Students should understand

- The purpose and usage of the renormalization group
- The connection between renormalization and scale setting
- Modern scale setting schemes

Resources

- DeGrand and DeTar, *Lattice Methods for Quantum Chromodynamics*, Chapter 4 (mostly 4.1)
- Leo Radzihovsky, Advanced Statistical Mechanics lecture notes, Lecture Set 5
- Rainer Sommer, *Scale setting in lattice QCD* [arXiv:1401.3270]
- FLAG 21, Section 11

1 Introduction

Suppose we understand some physical system at a deep theoretical level. In other words, we know the Lagrangian (or action) for this system. Typically, we want to use this theory to make predictions at some characteristic energy scale. We also know that this theory is incomplete in that there is some highest momentum (shortest length) scale Λ above which we cannot apply the theory (i.e., use it to do calculations). For this to be a useful (predictive) theory, Λ is presumably much higher than the energy scale of interest. This begs the question, how do we extract low energy physics from such a fundamental theory? Ideally, we would like to only work with the low energy degrees of freedom that are relevant to our application. But how do we identify which degrees of freedom these are, and how do we "thin out" the fundamental, high energy degrees of freedom to obtain a simpler, low energy effective theory?

The answers to these questions lie with the so-call "renormalization group" (RG), which gives us a systematic way to identify the relevant degrees of freedom and thin out the irrelevant ones. Conceptually, this amounts to coarse graining the system so that it is insensitive to the shortest length (highest momentum) modes. Analyzing the dynamics induced by RG allows us to tune our theoretical system so that we may match it onto the physical world. This framework is essential in determining the relationship between lattice spacing and coupling in QCD and gives a procedure for translating lattice results to physical predictions.

In Section 2, we define RG transformations and outline the key steps of how to use RG to extract physical results from theoretical systems. In Section 3, we see that this matching procedure amounts to tuning some number of parameters to match a set of experimentally determined quantities.

2 Renormalization

Here we describe the renormalization procedure. Roughly speaking, renormalization amounts to three main steps:

1. Systematically integrate out high energy degrees of freedom to obtain a low energy effective action. This procedure defines the renormalization group (RG) transformation.
2. Characterize/analyze the fixed points of the RG transformation. As we will see, the fixed points of the RG transformation correspond to the critical phenomena of the system.
3. Match physical observables (i.e., correlation functions) of the bare theory to nature at the RG fixed points.

2.1 The low energy effective action

A key step in renormalization is to obtain a low energy effective action. This will allow us to describe low energy physics irrespective of the high energy modes that our theory fails to describe. Here, we show two conceptually illustrative procedures for obtaining this low energy effective action and discuss their pros and cons.

2.1.1 Real-space RG

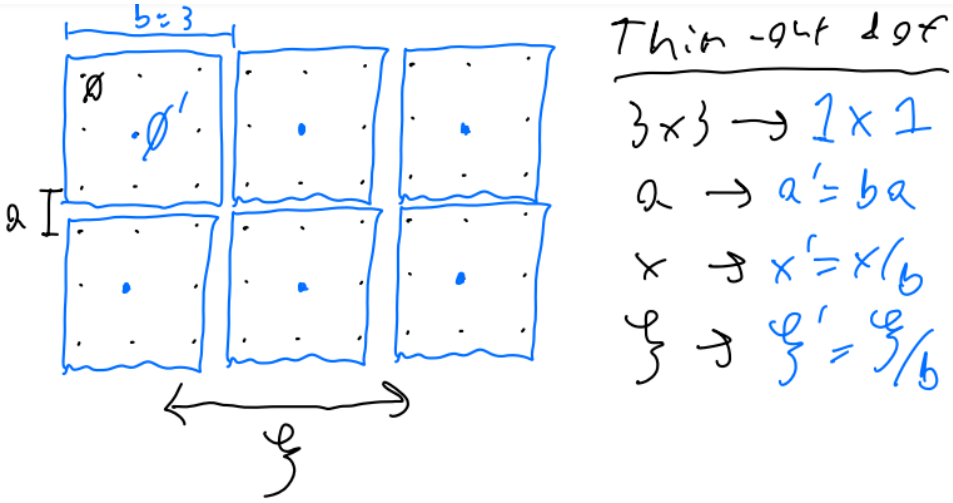

Figure 1: Real space RG blocking transformation. Adapted from Leo Radzihovsky's notes.

Consider a scalar field ϕ_x defined on discrete lattice points x and an action $S(\phi_x)$ in d dimensional Euclidean space. We will consider "thinning out" the short-scale (high energy) degrees of freedom by iteratively coarse-graining our lattice. This coarse graining will partition our lattice into blocks of side length b and volume $V = b^d$, effectively mapping our lattice with lattice spacing a to one with lattice spacing $a' = ba$ (see Fig. 2). There are a number of coarse graining procedures (e.g., decimation, majority rule). We will consider averaging our field ϕ_x over the volume V_x that is centered at site x. Hence, our fields transform as

$$\phi'_x = \frac{1}{b^d} \sum_{y \in V_x} \phi_y. \tag{1}$$

Note that we only apply this transformation at the new lattice sites x' situated at the centers of each block. This way the coarse graining is a partitioning and reduces the total number of degrees of freedom.

Let's consider how our partition function

$$Z = \int \mathcal{D}\phi_x \exp[-S(\phi)], \tag{2}$$

transforms under coarse graining. Note that

$$1 = \int \mathcal{D}\phi'_x \prod_{x'} \delta\left(\phi'_{x'} - \frac{1}{b^d}\sum_{y\in V_{x'}} \phi_y\right), \tag{3}$$

which is the statement that we only have one transformed field per block. Thus,

$$Z = \int \mathcal{D}\phi_x \mathcal{D}\phi'_x \prod_{x'} \delta\left(\phi'_{x'} - \frac{1}{b^d}\sum_{y\in V_{x'}} \phi_y\right) \exp[-S(\phi)] \tag{4}$$

$$= \int \mathcal{D}\phi'_{x'} \exp[-S'(\phi')], \tag{5}$$

where we have defined the transformed Boltzmann weights by integrating out the original fields

$$\exp[-S'(\phi')] \equiv \int \mathcal{D}\phi_x \prod_{x'} \delta\left(\phi'_{x'} - \frac{1}{b^d}\sum_{y\in V_{x'}} \phi_y\right) \exp[-S(\phi)]. \tag{6}$$

Eq. (6) defines the low energy effective action that results from a single coarse graining. While the real space blocking transformation provides some key intuition, Eq. (6) is clearly a cumbersome mathematical object to work with. Since the blocking procedure is discrete, it can be difficult to analyze the behaviour of the action under repeated transformations. Furthermore, there is no small parameter with which to analyze S' perturbatively. These practical issues motivate the definition of other schemes.

2.1.2 Momentum shell RG

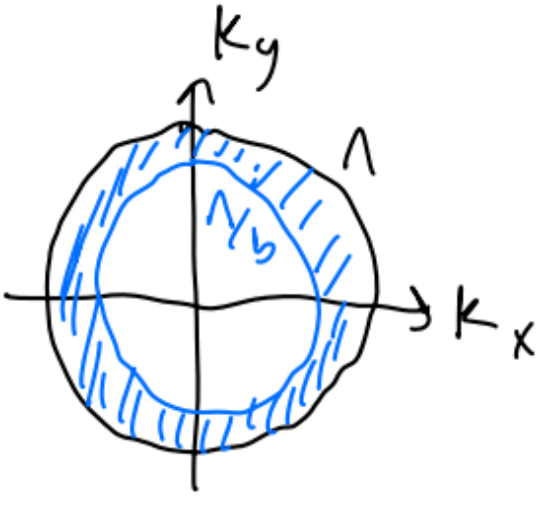

Figure 2: Momentum shell RG. Adapted from Leo Radzihovsky's notes.

As an alternative to the real-space blocking transformation defined in Section 2.1.1, we can perform an equivalent procedure in momentum space. In momentum space, the coarse graining that was defined by integrating over a volume in real space corresponds to integrating over a "shell" of momentum on the periphery of the Brillouin zone. The momentum shell to be integrated out is defined by $\Lambda' < k < \Lambda$ where $\Lambda = 1/a$ is the UV cutoff of the original lattice and $\Lambda' = \Lambda/b$ of the coarse grained lattice. Like before, our first goal will be to write the partition function in terms of the low energy fields with $0 < k < \Lambda'$ by integrating out the high energy degrees of freedom where now we identify "high energy" as "high momentum."

In anticipation of this integration, we can decompose our fields ϕ into their low and high frequency contributions

$$\phi = \phi_{\mathrm{L}} + \phi_{\mathrm{H}}, \tag{7}$$

where ϕ_{L} has $0 < k < \Lambda'$ and ϕ_{H} has $\Lambda' < k < \Lambda$. Hence,

$$\begin{aligned} Z &= \int \mathcal{D}\phi \exp[-S(\phi)] && (8)\\ &= \int \mathcal{D}\phi_{\mathrm{L}} \mathcal{D}\phi_{\mathrm{H}} \exp[-S(\phi_{\mathrm{L}}, \phi_{\mathrm{H}})] && (9)\\ &= \int \mathcal{D}\phi_{\mathrm{L}} \exp[-S'(\phi_{\mathrm{L}})], && (10) \end{aligned}$$

where

$$\exp[-S'(\phi_{\mathrm{L}})] \equiv \int \mathcal{D}\phi_{\mathrm{H}} \exp[-S(\phi_{\mathrm{L}}, \phi_{\mathrm{H}})]. \tag{11}$$

By comparing Eq. (11) to Eq. (6), we see the slight loss of conceptual clarity of momentum shell RG compared to real space RG affords us a simpler expression for the low energy effective action. A particularly nice feature of Eq. (11) is that it can be computed perturbatively.

2.2 The RG equation

Formally, we can write our low energy effective action as

$$S' = \mathcal{R}_b S, \tag{12}$$

where $\mathcal{R}_b$ is a nonlinear operator that transforms the action parameterized by the characteristic length scale of the coarse graining transformation b. This is the so-called "renormalization group equation" for the action.[1] Of course, we could perform a second coarse graining by b', in which case we expect

$$S'' = \mathcal{R}_{b'} S' = \mathcal{R}_{b'} \mathcal{R}_b S = \mathcal{R}_{b'b} S. \tag{13}$$

Typically, we are interested in actions of the form

$$S(\phi) = \sum_i g_i S_i(\phi), \tag{14}$$

[1] Of course, "group" is a bit of a misnomer—RG is a semi-group since R_b has no inverse.

where the S_i's are coordinate space integrals over field variables and their derivatives and the g_i's are the corresponding couplings. Under transformation by $\mathcal{R}_b$, we obtain

$$S'(\phi') = \sum_i g'_i S_i(\phi'). \tag{15}$$

In general, new operators can and will arise after blocking and are formally present in the starting action with zero coupling.[2]

It is often convenient[3] to also rescale our fields during the blocking transformation. Specifically, let

$$\phi'_{x'} = b^\zeta \Phi_{x=x'/b}, \tag{16}$$

where ζ is yet to be determined. Rescaling by a power of b ensures compositions of blockings scale appropriately. With these additional scalings, we can write

$$\sum_i g'_i S_i(\phi'_{x'}) = \sum_i g_{b,i} S_i(\Phi_x). \tag{17}$$

Note that we have assumed the S_i's are homogeneous in ϕ, as is typically the case. Things have gotten a bit technical and subtle, but the point of what we have just shown is that the transformed action can be written in the exact same form as the original action, meaning the fields are defined on every site (rather than just those leftover after blocking); this can be seen by relabeling Φ with ϕ. The purpose of this rewriting is that rather than worrying about renormalization of the action via $\mathcal{R}_b$, we can define an equivalent operator R_b that instead acts on the coupling constants

$$g_b = R_b(g). \tag{18}$$

So, we now also have the renormalization group equation for the couplings. Note that composition for R_b works as for $\mathcal{R}_b$, i.e., $R_{b'} R_b = R_{b'b}$.

Let's consider an example with a ϕ^4 scalar theory

$$S[\phi(x)] = \int \mathrm{d}^d x \left(\frac{1}{2} K (\partial_\mu \phi)^2 - \frac{1}{2} m \phi^2 + \lambda \phi^4 \right). \tag{19}$$

Normally we set the coupling on the kinetic term K to 1, which we will do shortly. In terms of the scaled fields and coordinates, the action is

$$S[\phi'(x')] = \int \mathrm{d}^d x' \left(\frac{1}{2} (b^{d-2+2\zeta} K) (\partial_{\mu'} \phi')^2 - \frac{1}{2} (b^{d+2\zeta} m) \phi'^2 + (b^{d+4\zeta} \lambda) \phi'^4 \right). \tag{20}$$

Therefore,

$$K_b = b^{d-2+2\zeta} K, \tag{21}$$

$$m_b = b^{d+2\zeta} m, \tag{22}$$

$$\lambda_b = b^{d+4\zeta} \lambda. \tag{23}$$

[2] This is why the S_i's don't need primes.

[3] This "convenience" effectively removes a redundant dimension in coupling space that results from the invariance of the partition function under multiplicative shifts of the Lagrangian. While not physically important, removing this redundancy can be important for guaranteeing the existence of RG fixed points in theory space. This is a common point of discussion in the gradient flow literature.

Note that we have not performed a blocking, merely performed dimensional analysis to get the scaling of the parameters. This sometimes referred to as "zeroth order RG." As seen from the above formulas, we can take K to not flow by setting

$$\zeta = (2-d)/2, \tag{24}$$

which gives

$$K_b = K, \tag{25}$$
$$m_b = b^2 m, \tag{26}$$
$$\lambda_b = b^{4-d}\lambda. \tag{27}$$

This is why we are free to set $K_b = K = 1$ at every scale. For emphasis, this is a choice made for convenience, not a requirement.

For first and higher order RG, we would set $\phi' = \phi_{\mathrm{L}} + \phi_{\mathrm{H}}$. The terms in the action would become

$$K(\partial_{\mu'}\phi')^2 = K[(\partial_{\mu'}\phi_{\mathrm{L}})^2 + 2(\partial_{\mu'}\phi_{\mathrm{L}})(\partial_{\mu'}\phi_{\mathrm{H}}) + (\partial_{\mu'}\phi_{\mathrm{H}})^2], \tag{28}$$
$$m\phi'^2 = m[\phi_{\mathrm{L}}^2 + 2\phi_{\mathrm{L}}\phi_{\mathrm{H}} + \phi_{\mathrm{H}}^2], \tag{29}$$
$$\lambda\phi'^4 = \lambda[\phi_{\mathrm{L}}^4 + 4\phi_{\mathrm{L}}^3\phi_{\mathrm{H}} + 6\phi_{\mathrm{L}}^2\phi_{\mathrm{H}}^2 + 4\phi_{\mathrm{L}}\phi_{\mathrm{H}}^3 + \phi_{\mathrm{H}}^4], \tag{30}$$

Plugging these into Eq. (11) gives an integral over ϕ_{H} that can be evaluated perturbatively. By matching to the original action (specifically about a RG fixed point to be discussed below), we can obtain higher order corrections to the coefficients. Note that $\phi_{\mathrm{H}} \to 0$ limit restores the zeroth order RG scaling above. These details won't be covered here, but are discussed at length in Leo Radzihovsky's notes.

2.2.1 RG fixed points

R_b as defined in Eq. (18) is generally a nonlinear operator. This affords us the use of the many tools developed in the field of nonlinear dynamics. Of particular interest here will be linear fixed point analysis.

To motivate the importance of fixed points, consider the simplest case where there is only one parameter g. In this case R_b can be defined by a scalar function f such that $g_b = f(g)$. Suppose f behaves as in Fig. 3. The sign of $f(g) - g$ determines the behavior of g under blocking: positive values mean g increases, negative g decreases. This means that if $g > g_1^*$, g will become infinite. For $g_2^* < g < g_1^*$, g will decrease and thus "flow" to g_2^*; similarly for $g < g_2^*$.

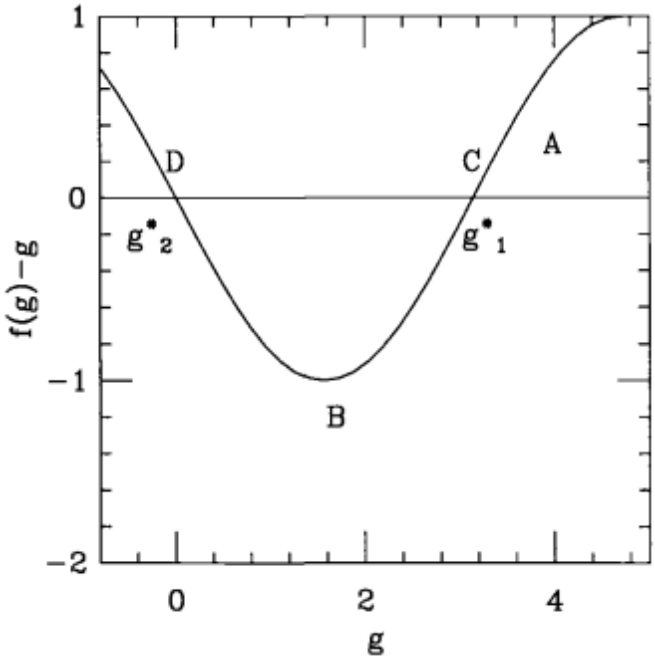

Figure 3: Single parameter RG flow example. Adapted from DeGrand and DeTar.

What this example emphasizes is that the long distance physics is fully determined by the fixed points of the RG transformation and the position of the bare parameters relative to these fixed points. For example, the long distance behavior is drastically different on either side of the g_1^* fixed point. Hence, to study the critical states of our theory, we must understand the fixed points of the RG transformation.

With the addition of more couplings, coupling constant space becomes more interesting. For example, Fig. 4 shows an example with two couplings. The fixed point has one attractive direction and one repulsive direction. If the bare theory has coupling constants anywhere on the attractive "critical line" then it blocks to the same fixed point theory. In higher dimensions, this generalizes to the "critical surface." The set of all theories that lie on the same critical surface comprise a so-called "universality class," which is uniquely determined by the dimensionality of spacetime and internal symmetries of the bare actions.

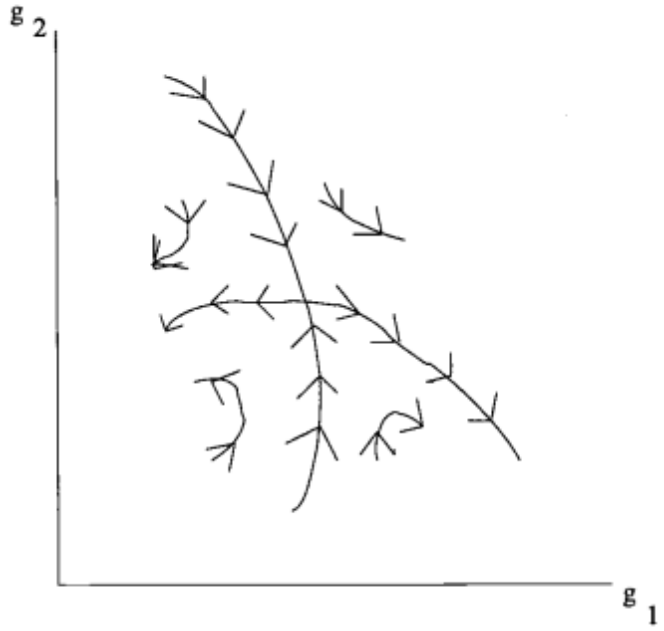

Figure 4: RG flow in a two-dimensional coupling space. Adapted from DeGrand and DeTar.

A simple yet illuminating way to categorize fixed points is linear stability analysis. Near a fixed point g^*, we can write $g = g^* + \delta g$ and hence "linearize" the RG equation $g_b = R_b(g)$ to obtain

$$\delta g_b = R_b^L \delta g + O(\delta g^2), \tag{31}$$

where

$$(R_b^L)_{\alpha\beta} = \left. \frac{\partial g_{b,\alpha}}{\partial g_\beta} \right|_{g=g^*}. \tag{32}$$

The eigenvectors e_j of R_b^L correspond to the directions in coupling space which the RG transform is a pure scale change. Since the eigenvalues λ_b must satisfy $\lambda_{b'}\lambda_b = \lambda_{b'b}$, we set $\lambda_j(b) = b^{y_j}$. Expanding

$$\delta g = \sum_j t_j e_j, \tag{33}$$

and blocking by b, we see

$$\delta g_b = \sum_j t_{b,j} e_j, \tag{34}$$

where $t_{b,j} = t_j b^{y_j}$. Hence, we can characterize the behavior near a fixed point according to the sign of the y_j's:

- $y_j > 0$: $t_{b,j}$ grows with b, and we move away from the fixed point under RG transformations. In this case, we call t_j a "relevant parameter" and the operator it multiplies a "relevant operator,"
- $y_j < 0$: $t_{b,j}$ vanishes under RG transformations and we approach the fixed point. These parameters/operators are called "irrelevant,"
- $y_j = 0$: R_s^L does not change the parameter/operator, which we call "marginal;" in the marginal case, higher order terms ($O(\delta g^2)$) dictate stability.

Note that the notion of "relevance" is only defined with respect to the fixed point, and may be different about other fixed points. The subspace of e_j's with $y_j < 0$ define the critical surface near g^*. See Fig. 5 for a visual of irrelevant and relevant operators.

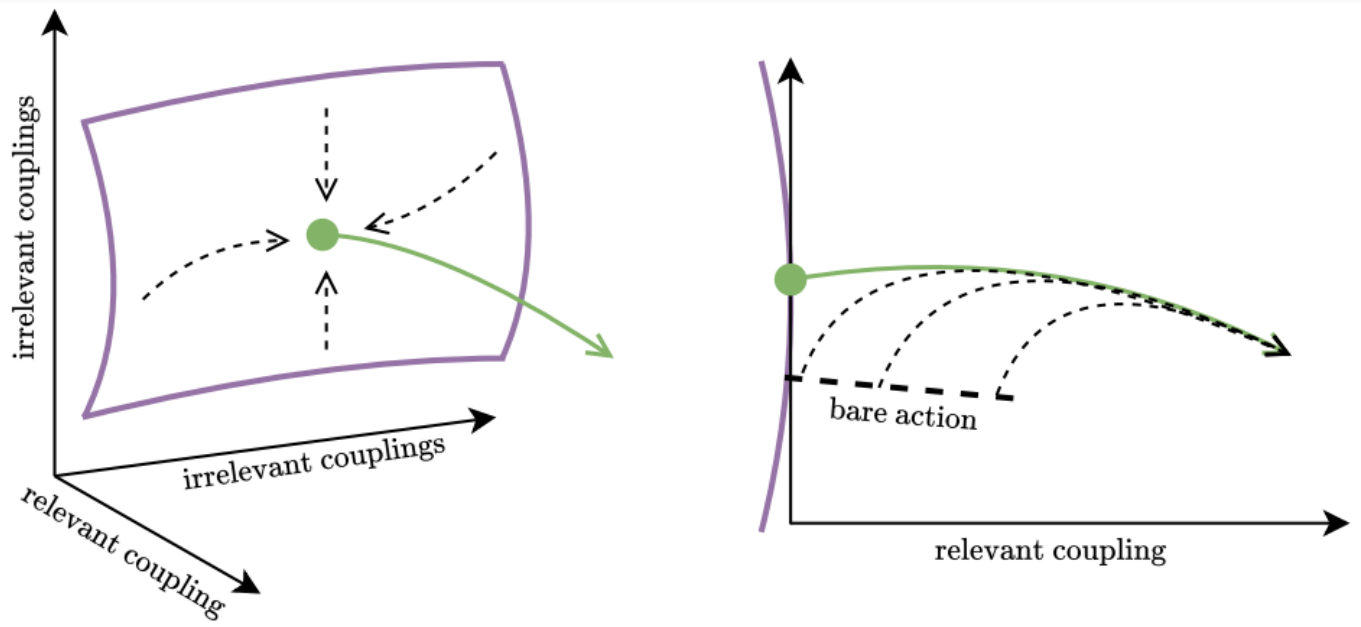

Figure 5: Illustration of RG flow from for a system with a single relevant parameter. Adapted from Curtis Peterson's thesis "Application of contemporary renormalization group techniques to strongly-coupled field theories."

Typically, the irrelevant operators flow to zero, and we need only concern ourselves with the relevant and marginal operators. The exceptional case is when an observable becomes singular as an irrelevant parameter flows to zero. We call this a "dangerously irrelevant parameter." Such parameters should be analyzed similarly to relevant and marginal parameters. See your favorite QFT textbook for more discussion.

2.3 Getting physics out of RG

Now let's consider how observables transform under blocking. Consider a two point correlator, which is often how we extract observables in lattice gauge theory. It is straightforward to see that

$$C(k) = \left\langle \phi_k^2 \right\rangle = b^{2\zeta+d} \left\langle {\phi'_{k'}}^2 \right\rangle . \tag{35}$$

The factors of b^ζ come from the wave function renormalization Eq. (16), and the remaining factors from the momentum integral. Therefore,

$$C(k, \{g\}) = b^{2\zeta+d} C_b(bk, \{g_b\}). \tag{36}$$

Here we see the power of RG. The physical correlator we are interested in is $C(k, \{g\})$, but this object is difficult to calculate directly as it is near the critical surface, and hence plagued with divergences related to the divergent correlation length ξ. However, we can relate this correlator (via RG) to $C_b(bk, \{g_b\})$, which can be calculated in a low energy effective theory. Of course, computing C_b and $\{g_b\}$ is non-trivial in general and requires perturbation theory or the definition of some non-perturbative scheme (this will be the subject of next week's lectures), but in principle there is no obstacle to doing to so.

2.3.1 Importance in lattice QCD

The bare action for an asymptotically free theory like QCD is characterized by one marginally relevant operator $F_{\mu\nu}^2$ with coupling g^2 (plus many irrelevant operators). In lattice, the bare actions are described by one overall factor and arbitrary weights of various closed loops

$$\beta S = \frac{2N}{g^2} \sum_j c_j S_j. \tag{37}$$

Asymptotic freedom means the critical surface of any RG transformation is at $g^2 = 0$. The location of the fixed point involves the relation of the c_j's. Setting g small but non-zero amounts to defining a bare theory that is slightly off of the critical surface. Under blocking, the theory then flows away from the critical surface along the renormalized trajectory (regardless of the values of the c_j's). This is how lattice QCD can be predictive. If some discretization of lattice QCD is in the same universality class as continuum QCD, then both theories are of the form Eq. (37) with different $\{c_j\}$. Hence, their long distance behavior is described by the same renormalized trajectory. This means we can simply work with the lattice theory where we know how to do nonperturbative (i.e., numerical) calculations.

Furthermore, the concept of irrelevant operators is important to improvement schemes on the lattice. In particular, we can add operators to our lattice action so that they cancel with discretization effects to some order in the lattice spacing a. So long as these additional operators are irrelevant, we still flow to the same line of constant physics and nothing about our RG procedure is changed.

2.4 Gradient flow

Now that we have set the stage general RG, it will be useful to discuss the related topic of gradient flow. For present purposes, gradient flow can be thought of as an alternative to real space RG via blocking transformations or momentum shell RG. Unlike those approaches, gradient flow is a continuous transformation that can be used to define and calculate renormalized quantities non-perturbatively, and has gained interest in the lattice community for these reasons. Here, we merely define gradient flow, and return to its usage in the scale setting section.

2.4.1 Definition and intuition

Consider a Euclidean scalar field theory with a single field $\phi(x)$. We define the gradient flow transformation by a diffusion equation:

$$\frac{\mathrm{d}\chi_t(x)}{\mathrm{d}t} = -\left.\frac{\delta S[\phi]}{\delta \phi}\right|_{\phi=\chi_t(x)}, \tag{38}$$

where S is the so-called flow action, t is the flow time, and the transformed field obeys the initial condition

$$\chi_{t=0}(x) = \phi(x). \tag{39}$$

Here, the flow time t can be thought of as the analog of b from the blocking transformation, the advantage here being that t is continuous variable. Note that t parameterizes flow along a renormalized trajectory and does not correspond to physical time; in fact, t has units of length (time) squared.

To get some intuition for how the gradient flow transformation is able to suppress high momentum modes, we specialize to the simplest case where S is a massless, free-field action. In this case, the gradient flow transformation reduces to the heat equation

$$\frac{\mathrm{d}\chi_t(x)}{\mathrm{d}t} = \nabla^2 \chi_t(x). \tag{40}$$

The solution is given by the integration of the initial field against a heat kernel K_t defined as

$$K_t(z) = \frac{e^{-z^2/(4t)}}{(4\pi t)^{d/2}}. \tag{41}$$

To be explicit

$$\chi_t(x) = (K_t\phi)(x) = \int_y K_t(x-y)\phi(y) = \int \mathrm{d}^d y \frac{e^{-y^2/(4t)}}{(4\pi t)^{d/2}}\phi(x+y). \tag{42}$$

As characteristic of solutions to the heat equation, integrating against the heat kernel quickly "diffuses" away the high momentum modes hence smoothing the transformed field for all positive flow times.

To connect the gradient flow transformation back to RG via the blocking transformation, consider the transformed field in some region of characteristic length scale $b \sim t^{1/2}$ centered at x_0. At long flow times, we deduce

$$\chi_t(x_0) = \int \mathrm{d}^d y \frac{e^{-(y/b)^2/4}}{(4\pi b^2)^{d/2}}\phi(x_0+y) \sim \int_{x\in V_{b(x_0)}} \frac{\mathrm{d}^d y}{|V_b(x_0)|}\phi(x_0+y), \tag{43}$$

where $V_b(x_0)$ is a d-dimensional volume (e.g., a ball of radius $b/2$) centered at x_0 with volume $|V_b(x_0)| \sim b^d$. The limits of integration change as the support is exponentially suppressed outside of $V_b(x_0)$. Identifying the right-hand side as the volume average of ϕ over $V_b(x_0)$, we see that the gradient flow transformation approaches the typical block-spin decimation procedure.[4]

Gradient flow can also be though of as a smoothed out version of momentum shell RG. Specifically, the hard momentum cutoff above which degrees of freedom are completely integrated out in momentum shell RG is replaced by a gradual transition from partially integrated degrees of freedom to almost entirely integrated degrees of freedom. See Fig. 6.

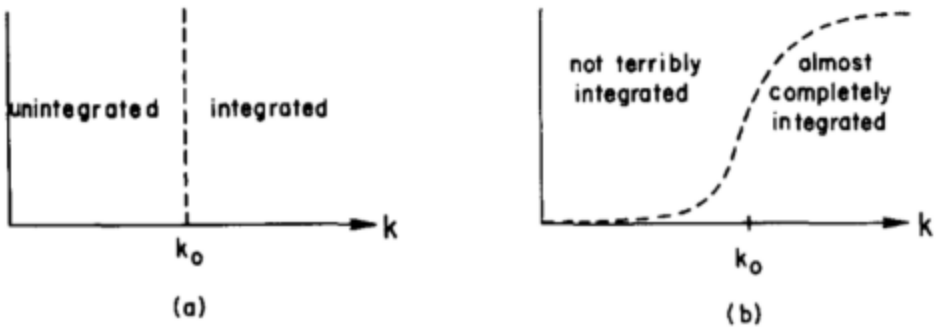

Figure 6: Qualitative distinction of sharp and smooth high momenta elimination under RG transformations. Adapted from Wilson and Kogut's "The renormalization group and the ε expansion" (1974).

2.4.2 Gradient flow for lattice gauge theories

In the continuum gauge theory, we defined the gradient flow transformation for the gauge fields $B_\mu(x,t)$ analogously to above:

$$\frac{\mathrm{d}B_\mu}{\mathrm{d}t} = -\left.\frac{\delta S[A]}{\delta A_\mu}\right|_{A_\mu = B_\mu(x,t)}, \qquad B_\mu(x,0) = A_\mu(x). \tag{44}$$

[4]To obtain the usual form of the coarse-graining transformation, we should also rescale the coordinates and fields. Since this just a heuristic argument to give some intuition, we omit these technical details for now.

Specializing to the Yang-Mills action $S_{\mathrm{YM}}[A]$, we deduce

$$\frac{\mathrm{d}B_\mu}{\mathrm{d}t} = D_\nu G_{\nu\mu}, \tag{45}$$

where

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + [B_\mu, B_\nu]\,, \tag{46}$$

is the flowed field strength tensor and

$$D_\mu = \partial_\mu + [B_\mu, \cdot], \tag{47}$$

is the flowed covariant derivative.

On the lattice, the gradient flow transformation is defined analogously except using the Wilson action $S_{\mathrm{W}}[U]$. Let $V_\mu(x,t)$ denote the flowed link variable corresponding to $U_\mu(x)$. The flow transformation is defined as

$$\frac{\mathrm{d}V_\mu(x,t)}{\mathrm{d}t} = -g_0^2\{\partial_\mu S_{\mathrm{W}}[V_t]\}V_\mu(x,t) \tag{48}$$

$$= -g_0^2 T^a \partial_\mu^a S_{\mathrm{W}}[V_t]V_\mu(x,t), \tag{49}$$

where T^a are the generators of $\mathrm{SU}(N)$. The spatial derivative is Lie algebra-valued and defined as

$$\partial_\mu^a f[U] = \frac{\mathrm{d}}{\mathrm{d}s} f\left[e^{sX_\mu(x)U}\right]\Big|_{s=0}, \tag{50}$$

where $X_\mu(x) = \delta_{\mu\nu}\delta(x,y)T^a$ and $U \equiv \{U_\nu(y)\}$ denote the set of gauge links on the lattice. See [arXiv:1006.4518] for more details on these definitions and gradient flow on the lattice.

With the gradient flow defined, there is no obstacle in principle to the rest of the renormalization procedure outlined above. In practice, this done numerically. Gradient flow is often less computationally expensive than other non-perturbative RG schemes (e.g., Monte Carlo RG).

2.5 Exercises

1. Compare real space RG via blocking to momentum shell RG. Specifically, verify that the momentum space fields ϕ_k approximately match the Fourier transform of the real space fields ϕ_x. What is the momentum space analog of the real space "wave function renormalization" condition defined in Eq. (16)? Hint: See DeGrand and DeTar for guidance.

2. Problem 4.2 from Mehran Kadar's "Statistical Physics of Fields." Start with parts (a) and (b), and only do the others if you have time and are interested. Note that the given differential equations define the linearized RG transformation of the Ising model in $d = 1 + \epsilon$ dimensions with scaling parameter $b = e^\ell$. In our notation, the bare parameters g are T and h.

2. *The Ising model:* The differential recursion relations for temperature T, and magnetic field h, of the Ising model in $d = 1+\epsilon$ dimensions are (for $\mathrm{b} = \mathrm{e}^{\ell}$)

$$\begin{cases} \dfrac{\mathrm{d}T}{\mathrm{d}\ell} = -\epsilon\, T + \frac{1}{2}T^2 \\ \dfrac{\mathrm{d}h}{\mathrm{d}\ell} = dh. \end{cases}$$

(a) Sketch the renormalization group flows in the (T, h) plane (for $\epsilon > 0$), marking the fixed points along the $h = 0$ axis.

(b) Calculate the eigenvalues y_t and y_h, at the critical fixed point, to order of ϵ.

(c) Starting from the relation governing the change of the correlation length ξ under renormalization, show that $\xi(t, h) = t^{-\nu} g_\xi\left(h/|t|^{\Delta}\right)$ (where $t = T/T_c - 1$), and find the exponents ν and Δ.

(d) Use a hyperscaling relation to find the singular part of the free energy $f_{\text{sing}}(t, h)$, and hence the heat capacity exponent α.

(e) Find the exponents β and γ for the singular behaviors of the magnetization and susceptibility, respectively.

(f) Starting with the relation between susceptibility and correlations of local magnetizations, calculate the exponent η for the critical correlations ($\langle m(\mathbf{0})m(\mathbf{x})\rangle \sim |\mathbf{x}|^{-(d-2+\eta)}$).

(g) How does the correlation length diverge as $T \to 0$ (along $h = 0$) for $d = 1$?

3. Show that to lowest order in the weak coupling expansion, the flow equation Eq. (44) for the Yang-Mills actions $S_{\mathrm{YM}}[A]$ becomes the heat equation and thus

$$B_\mu(x,t) = \int \mathrm{d}^4 y \frac{e^{-(x-y)^2/(4t)}}{(4\pi t)^{1/2}} A_\mu(y) + O(g_0^2). \tag{51}$$

3 Scale setting

Scale setting is an important step required to make physical predictions from lattice simulations. As we will see, scale setting amounts to fixing the quark masses and matching an overall scale to experiment. While not a very practical choice, it is useful to think of the overall scale as something like a nucleon mass to provide intuition as we go along. Many different scale setting schemes have been developed. In principle, results should have no dependence on the choice of scheme; however, in practice the statistical and systematic precision on physical inputs as well as the lattice measurements to which these inputs are matched lead to different pros and cons for each scheme.

The uncertainty from lattice analyses are rather sensitive to uncertainties in the scale.[5] Denoting $\mathcal{S}$ as the overall scale, one finds that the scale uncertainty $\delta\mathcal{S}$ propagates about linearly on to that of hadron and quark mass M and m, respectively,

$$\frac{\delta M}{M} \sim \frac{\delta m}{m} \sim \frac{\delta \mathcal{S}}{\mathcal{S}}. \tag{52}$$

This can be even worse for other observables. An example of much current interest is the hadronic vacuum polarization contribution to the anomalous magnetic moment of the muon (also known as muon $g-2$), denoted a_μ^{HVP}, which is currently being determined on the lattice to the per mille level. In this case, it can be shown that the error propagation is approximately quadratic

$$\frac{\delta a_\mu^{\mathrm{HVP}}}{a_\mu^{\mathrm{HVP}}} \sim 2\frac{\delta \mathcal{S}}{\mathcal{S}}. \tag{53}$$

This means the scale must be known to the level of a few per mille for continued progress on muon $g-2$ from the lattice.

With the importance of scale in mind, let's think about what makes a good scale. Generally, there are four key features:

1. Inexpensive to compute,
2. Good statistical precision,
3. Good systematic precision,
4. Weak quark mass dependence.

The first two are self-explanatory. Recalling that the overall scale is something like a nucleon mass, the systematic effects that concern point 3 include finite volume effects, excited state contamination, and discretization error. Point 4 makes extrapolation/interpolation of simulations not at the physical point easier when the scale is insensitive to the tuning of the bare quark masses.

It is useful to distinguish between "physical scales" (also known as phenomenological scales) and "theory scales." The former are physical quantities (something like a nucleon mass) that are observable with minimal theoretical considerations; the latter are constructed to be easily and precisely determined through lattice QCD. Currently, theory scales serve as proxies, each relying on a physical scale as input. In practice, this means that the continuum limit of a theory scale needs to be taken in units of a physical scale in order to connect to nature. Theory scales are useful because they are practical to use for scale setting and are designed to satisfy criteria 1-4 above more easily than the physical scale to which they are related.

3.1 What does scale setting have to do with RG?

Before moving on to specific scale setting schemes, it is worth discussing why this topic is covered in the renormalization unit of the course. In QFT, the need for renormalization becomes apparent when attempting to make predictions in absolute terms from theory alone; it is only after reference to physics at another (arbitrary) scale that we can make predictions about observables. From this perspective, we see the relevance of having some physical inputs with which to establish this reference scale.

[5] The examples of the relationship between observable uncertainty and scale uncertainty presented here come from combination of dimensional analysis-based arguments and numerical observations. See FLAG 11.1 for details.

Let's illustrate this for the case of lattice QCD. In QCD, the theory is completely specified with the N_f quark masses and the coupling strength g.[6] In principle, we could set the bare quark masses and coupling in a lattice theory to their corresponding physical values. However, the quark masses are not accessible experimentally without the use of perturbation theory, which only applies at short distances. Since lattice QCD is most effective at long distances (i.e., the limit of small lattice spacing), this is clearly not an effective way to proceed.

Instead, we would like to fix the N_f+1 parameters to non-perturbative, long-distance observables found in nature. An obvious choice would be N_f+1 hadron masses, preferably ones that are stable in QCD+QED (i.e., ones that only decay through electroweak processes, but not through strong processes). This leads to the hadronic renormalization scheme defined by

$$\frac{M_i(g_0, \{am_{0,j}\})}{M_1(g_0, \{am_{0,j}\})} = \frac{M_i^{\text{expt}}}{M_1^{\text{expt}}}, \qquad i = 2, \ldots, N_f+1, \qquad j = 1, \ldots, N_f, \tag{54}$$

where M_i are the chosen hadron masses, g_0 the bare coupling, and $am_{0,j}$ the bare quark masses in lattice units.[7] Under RG, $am_{0,j}$ flow toward a renormalized trajectory parameterized by the relevant parameter g_0. Tuning g_0 closer to the critical surface makes convergence to this renormalized trajectory quicker, and hence the lattice theory closer to the physical point. In other words, the continuum limit is achieved by taking $g_0 \to 0$, and Eq. (54) tells us how to tune our lattice observables corresponding to physical observables in this limit. More formally, consider the observable $\mathcal{O}$ with mass dimension $d_{\mathcal{O}}$. Defining the dimensionless ratio

$$\hat{\mathcal{O}}(aM_1) = \left.\frac{\mathcal{O}}{M_1^{d_{\mathcal{O}}}}\right|_{am_{0,j}\to \text{RT}_j(g_0)}, \tag{55}$$

where $\text{RT}_j(g_0)$ is the aforementioned renormalized trajectory. Continuum predictions are obtained as

$$\mathcal{O}^{\text{cont}} = (M_1^{\text{expt}})^{d_{\mathcal{O}}} \lim_{aM_1\to 0} \hat{\mathcal{O}}(aM_1). \tag{56}$$

With this notation, we see that the bare parameters $am_{0,j}$ (which come from the M_i's) are used to determine the dimensionless ratio $\hat{\mathcal{O}}$, and M_1 determines the overall scale. This is why determining aM_1 is what we mean by scale setting.

The question remains of what hadron masses should we use. Obtaining $am_{0,j} \to \text{RT}_j(g_0)$ must be done numerically, so it would be ideal if the i-th mass ratio depended predominantly on the i-th quark mass in order to decouple these numerical tasks. This motivates the use of pseudoscalar mesons for the M_i's where one can establish, to leading order, a one-to-one correspondence between m_{ud}, m_s, m_c, m_b and M_π, M_K, M_D, M_B through, for example, chiral perturbation theory.[8] To leading order, the Gell-Mann–Oakes–Renner (GMOR) relation tells us that the square of the pseudoscalar meson masses are proportional the sum of their constituent quark masses, i.e.,

$$M_{q_1q_2}^2 = B_0(m_{q_1} + m_{q_2}). \tag{57}$$

This leading order result works well for the light quarks (pions) and reasonably well with strange quarks (kaons). For heavier flavors, the relationship becomes approximately linear, i.e., $M_D \propto m_c$ and $M_B \propto m_b$.

Choosing M_1 is more difficult and leads to the various schemes described below.

[6] Here we ignore the possibility of CP violation introduced via the θ parameter.

[7] There are some caveats here with regard to finite volume corrections, see FLAG 11.2 for details. We omit these technicalities for simplicity, as they are not essential to the illustration, and assume the infinite volume limit has already been taken.

[8] Here we've simplified to the isospin-symmetric limit $m_{ud} = (m_u + m_d)/2$ and ignored the top quark.

3.2 Physical (phenomenological) scales

On physical grounds, a natural choice for M_1 would be the proton mass M_p. However, nucleon correlators are notoriously noisy, leading to large statistical uncertainties in M_p.[9] This has prompted the search for an alternative of which the most popular are the Omega baryon mass M_Ω and psuedoscalar decay constants f_π and f_K.

3.2.1 Omega baryon mass $\mathbf{M_\Omega}$

A natural alternative to M_p is another baryon mass. Considerations of stability and light quark mass dependence leads to the Omega baryon. M_ω has much better noise-to-single (N/S) properties compared to M_p. Specifically,

$$R^p_{N/S} \overset{x_0 \text{ large}}{\sim} K_p \exp[(M_p - \frac{3}{2}M_\pi)x_0] \overset{\text{PDG}}{\approx} \exp(x_0/0.27 \text{ fm}), \tag{58}$$

$$R^\Omega_{N/S} \overset{x_0 \text{ large}}{\sim} K_\Omega \exp[(M_\Omega - \frac{3}{2}M_{\eta_s})x_0] \overset{\text{PDG}}{\approx} \exp(x_0/0.31 \text{ fm}), \tag{59}$$

where $M^2_{\eta_s} \approx 2M^2_K - M^2_\pi$. While this my seem like a small difference at first glance, it is the difference between having a reasonable size plateau in M_Ω fits versus basically no discernible plateau in M_p fits (see Fig. 7). Currently, lattice groups are able to determine M_Ω to the sub-per mille level.

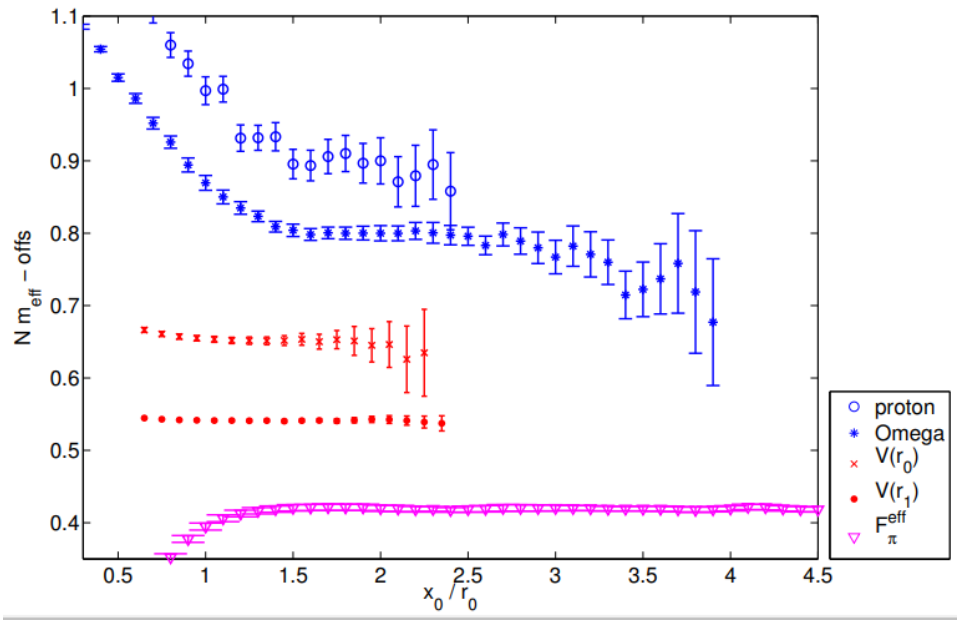

Figure 7: Example effective mass plots for M_p, M_Ω, $V(\approx r_0)$, $V(\approx r_1)$, and f_π from the CLS collaboration. Adapted from Rainer Sommer's "Scale setting in lattice QCD."

The other advantage of M_Ω is the weak dependence on the light quark masses (in the case of fixed strange mass). On the other hand M_Ω depends more on the strange mass than other schemes.

3.2.2 Psuedoscalar decay constants $\mathbf{f_\pi, f_K}$

Two other options that satisfy the 4 criteria outlined above are the pion and kaon decay constants f_π, f_K. A clear advantage from the perspective of lattice is that the pseudoscalar correlators are

[9] For an accessible review on the signal-to-noise ratio problem for the proton, see Peter Lepage's 1989 TASI lecture notes entitled "The Analysis of Algorithms for Lattice Field Theory."

very stable and, in fact, almost t-independent for sufficiently large t, giving long plateaus and high precision fits.

The drawbacks come from their experimental determination. The dominant contributions from weak decays processes. The $\pi \to \ell\nu$ and $K \to \ell\nu$ decay rates give products of $V_{ud}f_\pi$ and $V_{us}f_K$, respectively. Hence, the precision of these measurements is limited by our knowledge of the CKM matrix elements V_{ud}, V_{us}.

To get a sense of current levels of precision, some current FLAG averages with $N_f = 2 + 1$ dynamical fermion flavors are

$$|V_{ud}| = 0.97438(12), \qquad f_{\pi^+} = 130.2(8)\ \text{MeV}, \tag{60}$$
$$|V_{us}| = 0.2249(5), \qquad f_{K^+} = 155.7(7)\ \text{MeV}, \tag{61}$$

which correspond to around 6 per mille error on f_{π^+} and 4 per mille on on f_{K^+}.

3.3 Theory scales

As explained above, theory scales are used to enhance the precision of physical scales on the lattice. They serve as proxies for the overall scale M_1 in the limit used to obtain $\mathcal{O}^{\text{cont}}$ from Eq. (56). To see how this works consider the theory scale $\mathcal{S}$. We can compute $\mathcal{O}^{\text{cont}}$ using

$$\mathcal{O}^{\text{cont}} = (\mathcal{S}^{\text{phys}})^{d_\mathcal{O}} \lim_{a\mathcal{S}\to 0} \hat{\mathcal{O}}_\mathcal{S}(a\mathcal{S}), \tag{62}$$

where

$$\hat{\mathcal{O}}_\mathcal{S}(a\mathcal{S}) = \left.\frac{\mathcal{O}}{\mathcal{S}^{d_\mathcal{O}}}\right|_{am_{0,j}\to \text{RT}_j(g_0)}, \tag{63}$$

and the physical scale M_1 enters via

$$\mathcal{S}^{\text{phys}} = (M_1^{\text{expt}}) \lim_{aM_1\to 0} \hat{\mathcal{S}}_{M_1}(aM_1). \tag{64}$$

Note that the extrapolation in Eq. (64) that determines the overall theory scale from the physical scale only needs to be performed once, and thus the same theory scale can be used for many different analyses. Once the theory scale is determined, the continuum limit is taken in terms of the theory scale using Eq. (62) in subsequent analyses.

The most popular theory scales are those coming from the static quark potential r_0, r_1 and those from gradient flow t_0, w_0.

3.4 Scales from the static quark potential $\mathbf{r_0}, \mathbf{r_1}$

The static quark potential was discussed in lecture 8 by Ryan Abbott, the relevant part to us here being the definition of the Sommer parameter r_0 (see also 3.5.2 of Gattringer and Lang). In the interest of not being too repetive, I will keep the discussion of these scales brief.

To recap, the static quark potential is of the form

$$V(r) = A + \frac{\alpha}{r} + \sigma r, \tag{65}$$

where A is an arbitrary constant, α is the EM coupling strength, and σ is the string tension. Hence, the associated forces is

$$F(r) = \frac{\mathrm{d}V}{\mathrm{d}r} = -\frac{\alpha}{r^2} + \sigma. \tag{66}$$

The theory scale of interest is r_c such that

$$r^2 F(r)\big|_{r=r_c} = c. \tag{67}$$

For $c = 1.65$, we get the Sommer parameter $r_0 \equiv r_{1.65}$. Another proposed choice is r_1 for $c = 1$, which has some improved statistical precision due to the quicker decay of excited state contamination at smaller r.

Phenomenological arguments based on $b\bar{b}$, $c\bar{c}$ spectra give $r_0 \approx 0.49$ fm. While this may motivate the use of r_0 as a physical scale, the connection between phenomenological potentials and the static quark potential has not been made quantitative, and hence a physical scale is needed as well.

Current FLAG averages with $N_f = 2 + 1 + 1$ are

$$r_0 = 0.474(14) \text{ fm}, \qquad (\sim 300‰ \text{ error}) \tag{68}$$

$$r_1 = 0.3112(30) \text{ fm}. \qquad (\sim 10‰ \text{ error}) \tag{69}$$

3.5 Gradient flow scales $\mathbf{t_0}, \mathbf{w_0}$

The other scales discussed here are defined using gradient flow, which was defined in Section 2.4. Recall for gradient flow on the Yang-Mills action gives

$$\frac{\mathrm{d}B_\mu}{\mathrm{d}t} = D_\nu G_{\nu\mu}, \tag{70}$$

where

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + [B_\mu, B_\nu], \tag{71}$$

$$D_\mu = \partial_\mu + [B_\mu, \cdot]. \tag{72}$$

Furthermore, recall that flowed correlation functions are smooth and finite for any $t > 0$. In particular, we define the

$$\mathcal{E}(t) = t^2 \langle E(x,t) \rangle, \tag{73}$$

where

$$E(x,t) = -\frac{1}{2} \mathrm{Tr}[G_{\mu\nu}(x,t) G_{\mu\nu}(x,t)], \tag{74}$$

is the Yang-Mills energy density at finite flow time, which needs no renormalization beyond that of bare couplings and quark masses.

From this object we define two scales

$$\mathcal{E}(t_0) = 0.3, \tag{75}$$

$$w_0^2 \mathcal{E}'(w_0^2) = 0.3, \tag{76}$$

where $\mathcal{E}'(t) = \mathrm{d}\mathcal{E}(t)/\,\mathrm{d}t$. These are truly theory scales without phenomenological analogs (unlike r_0). This means they can only be computed via lattice simulation in terms of a physical scale and taking the continuum limit.

t_0 was proposed in [arXiv:1006.4518] by Martin Lüuscher and w_0 in [arXiv:1203.4469] by the BMW collaboration. The "more natural" of the two t_0 is motivated by the fact that the object $\mathcal{E}(t)$ is approximately linear in the flow time t, see Fig. 8. The other scale w_0 was inspired by the observation that this linear dependence was approximately parallel at many different lattice

spacings in the case of an M_Ω-based extrapolation as seen in Fig. 8. Hence, taking the derivative reduces discretization effects. The squares give w_0 units of length (as opposed to length squared for t_0). Using 0.3 for the right-hand sides of Eqs. (75) and (76) is another numerically motivated but arbitrary choice, and others have been considered.

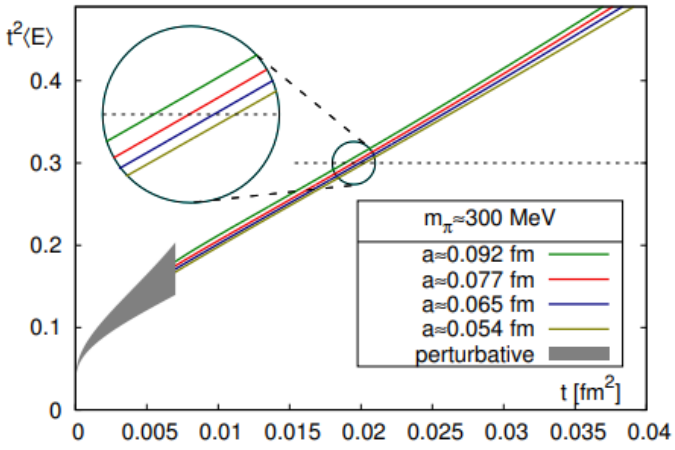

Figure 8: Continuum extrapolation of $t^2\mathcal{E}(t)$ from the BMW collaboration. Adapted from [arXiv:1203.4469].

Current FLAG averages with $N_f = 2+1+1$ are

$$\sqrt{t_0} = 0.14292(104) \text{ fm}, \qquad (\sim 7‰ \text{ error}) \tag{77}$$
$$w_0 = 0.17256(103) \text{ fm}. \qquad (\sim 6‰ \text{ error}) \tag{78}$$

Of the theory scales discussed, w_0 has the lowest relative error. FLAG also reports a $N_f = 1+1+1+1$ determination of $w_0 = 0.17236(70)$ fm, which achieves around 4 per mille error.

3.6 Exercises

1. Consider the so-called "pure-QCD world" pseudoscalar meson masses with $N_f = 1+1+1$ flavors

$$M_{\pi^+} = 135.0, \tag{79}$$
$$M_{K^+} = 491.4, \tag{80}$$
$$M_{K^0} = 497.6, \tag{81}$$

which defines the pseudoscalar meson masses without QED effects.

   (a) Using leading order GMOR, determine the set of pseudeoscalar mesons masses that define the strong isospin symmetric pure-QCD world with $N_f = 2+1$ flavors (i.e., in the case of a single average light quark mass $m_{ud} = (m_u + m_d)/2$). Hint: The isospin symmetric world will be defined by two mesons—a pion made of two light quarks, and a kaon made of a light quark and a strange quark.

   (b) In precision lattice calculations, it is common to compute the quantity $\Delta M_\pi^2 \equiv M_{\pi^{dd}}^2 - M_{\pi^{uu}}^2$, where π^{dd} (π^{uu}) is the pseudoscalar meson made of two down (up) quarks, to quantify strong isospin breaking effects. First, compute ΔM_π^2 in pure-QCD world using GMOR (Hint: The answer will only depend on the kaon masses). Next, redo the calculation using

the value of m_u/m_d from the PDG (with M_{π^+} still from pure-QCD world) instead of the kaon masses. Do the two results agree? Discuss with your groups the reasons why they do or do not.

2. Show that in the weak coupling limit,

$$\mathcal{E}(t) \sim g_0^2 t^2 \int_0^\infty \mathrm{d}p p^3 e^{-2tp^2} + O(g_0^4), \tag{82}$$

$$t\mathcal{E}'(t) \sim g_0^2 t^2 \int_0^\infty \mathrm{d}p p^3 (1 - tp^2) e^{-2tp^2} + O(g_0^4). \tag{83}$$

(Hint: Consider the Fourier transform of Eq. (51).) Plot the integrands as a function of p at $t = 0$ and some increasing values of t. Use this to argue that $\mathcal{E}$ is now finite with suppressed cutoff effects. In units of g_0^2 (i.e., setting $g_0 = 1$ for this part of the exercise), plot $\mathcal{E}(t)$ and $t\mathcal{E}'(t)$ as a function of t to leading order in g_0, and compute the weak coupling limit of the scales t_0 and w_0.