5 The Renormalization Group

Even a humble glass of pure water consists of countless H_2O molecules, which are made from atoms that involve many electrons perpetually executing complicated orbits around a dense nucleus, the nucleus itself is a seething mass of protons and neutrons glued together by pion exchange, these hadrons are made from the complicated and still poorly understood quarks and gluons which themselves maybe all we can make out of tiny vibrations of some string, or modes of a theory yet undreamed of. How then is it possible to understand anything about water without first solving all the deep mysteries of Quantum Gravity?

In classical physics the explanation is really an aspect of the Principle of Least Action: if it costs a great deal of energy to excite a degree of freedom of some system, either by raising it up its potential or by allowing it to whizz around rapidly in space–time, then the least action configuration will be when that degree of freedom is in its ground state. The corresponding field will be constant and at a minimum of the potential. This constant is the zero mode of the field, and plays the role of a Lagrange multiplier for the remaining low–energy degrees of freedom. You used Lagrange multipliers in mechanics to confine wooden beads to steel hoops. This is a good description at low energies, but my sledgehammer can excite degrees of freedom in the hoop that your Lagrange multiplier doesn't reach.

We must re-examine this question in QFT because we're no longer constrained to sit at an extremum of the action. The danger is already apparent in perturbation theory, for even in a process where all external momenta are small, momentum conservation at each vertex still allows for very high momenta to circulate around the loop and the value of these loop integrals would seem to depend on all the details of the high–energy theory. The Renormalization Group (RG), via the concept of *universality*, will emerge as our quantum understanding of why it is possible to understand physics at all.

5.1 Integrating out degrees of freedom

Suppose our QFT is governed by the action

$$S_{\Lambda_0}[\varphi] = \int \mathrm{d}^d x \left[\frac{1}{2}\partial^\mu\varphi\,\partial_\mu\varphi + \sum_i \Lambda_0^{d-d_i} g_{i0}\,\mathcal{O}_i(x)\right] . \tag{5.1}$$

Here we've allowed arbitrary local operators $\mathcal{O}_i(x)$ of dimension $d_i > 0$ to appear in the action; each $\mathcal{O}_i$ can be a Lorentz–invariant monomial involving some number n_i powers of fields and their derivatives, *e.g.* $\mathcal{O}_i \sim (\partial\varphi)^{r_i}\varphi^{s_i}$ with $r_i + s_i = n_i$. For later convenience, I've included explicit factors of some energy scale Λ_0 in the couplings, chosen so as to ensure that the coupling constants g_{i0} themselves are dimensionless, but of course the action is at this point totally general. We've simply allowed all possible terms we can include to appear.

Given this action, we can define a regularized partition function by

$$\mathcal{Z}_{\Lambda_0}(g_{i0}) = \int_{C^\infty(M)_{\leq\Lambda_0}} \mathcal{D}\varphi\; \mathrm{e}^{-S_{\Lambda_0}[\varphi]/\hbar} \tag{5.2}$$

where the integral is taken over the space $C^\infty(M)_{\leq\Lambda_0}$ of smooth functions on M whose energy is at most Λ_0. The first thing to note about this integral is that it makes sense:

we've explicitly regularized the theory by declaring that we are only allowing momentum modes up to the cut–off[19] Λ_0. For example, there can be no UV divergences[20] in any perturbative loop integral following from (5.2), because the UV region is simply absent.

Now let's think what happens as we try to perform the path integral by first integrating those modes with energy between Λ_0 and $\Lambda < \Lambda_0$. The space $C^\infty(M)_{\leq\Lambda_0}$ is naturally a vector space with addition just being pointwise addition on M. Therefore we can split a general field $\varphi(x)$ as

$$\begin{aligned}\varphi(x) &= \int_{|p|\leq\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d}\, \mathrm{e}^{\mathrm{i}p\cdot x}\, \tilde{\varphi}(p) \\ &= \int_{|p|\leq\Lambda} \frac{\mathrm{d}^d p}{(2\pi)^d}\, \mathrm{e}^{\mathrm{i}p\cdot x}\, \tilde{\varphi}(p) + \int_{\Lambda<|p|\leq\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d}\, \mathrm{e}^{\mathrm{i}p\cdot x}\, \tilde{\varphi}(p) \\ &=: \phi(x) + \chi(x)\,,\end{aligned} \tag{5.3}$$

where $\phi \in C^\infty(M)_{\leq\Lambda}$ is the low–energy part of the field, while $\chi \in C^\infty(M)_{(\Lambda,\Lambda_0]}$ has high energy. The path integral measure on $C^\infty(M)_{\leq\Lambda_0}$ likewise factorizes as

$$\mathcal{D}\varphi = \mathcal{D}\phi\ \mathcal{D}\chi$$

into a product of measures over the low– and high–energy modes. Performing the integral over the high–energy modes χ provides us with an *effective action at scale* Λ

$$S^{\text{eff}}_{\Lambda}[\phi] := -\hbar \log\left[\int_{C^\infty(M)_{(\Lambda,\Lambda_0]}} \mathcal{D}\chi\ \exp\left(-S_{\Lambda_0}[\phi+\chi]/\hbar\right)\right] \tag{5.4}$$

involving the low–energy modes only. We call the process of integrating out modes *changing the scale* of the theory. We can iterate this process, integrating out further modes and obtaining a new effective action

$$S^{\text{eff}}_{\Lambda'}[\phi] := -\hbar \log\left[\int_{C^\infty(M)_{(\Lambda',\Lambda]}} \mathcal{D}\chi\ \exp\left(-S^{\text{eff}}_{\Lambda}[\phi+\chi]/\hbar\right)\right] \tag{5.5}$$

at a still lower scale $\Lambda' < \Lambda$. For this reason, equation (5.4) is known as the *renormalization group equation* for the effective action.

Separating out the kinetic part, we write the original action as

$$S_{\Lambda_0}[\phi+\chi] = S^0[\phi] + S^0[\chi] + S^{\text{int}}_{\Lambda_0}[\phi,\chi] \tag{5.6}$$

where $S^0[\chi]$ is the kinetic term

$$S^0[\chi] = \int \mathrm{d}^d x \left[\frac{1}{2}(\partial\chi)^2 + \frac{1}{2}m^2\chi^2\right] \tag{5.7}$$

[19]In writing S_{Λ_0} in terms of dimensionless couplings, we used the same energy scale Λ_0 as we chose for the cut-off. This was purely for convenience.

[20]On a non–compact space–time manifold M there can be IR divergences. This is a separate issue, unrelated to renormalization, that we'll handle later if I get time. If you're worried, think of the theory as living in a large box of side L with either periodic or reflecting boundary conditions on all fields. Momentum is then quantized in units of $2\pi/L$, so the space $C^\infty(M)_{\leq\Lambda_0}$ is finite–dimensional.

for χ and $S^0[\phi]$ is similar. Notice that the quadratic terms can contain no cross–terms $\sim \phi\chi$, because these modes have different support in momentum space. For the same reason, the terms in the effective interaction $S^{\rm int}_{\Lambda_0}[\phi,\chi]$ must be at least cubic in the fields. Since ϕ is non–dynamical as far as the χ path integral goes, we can bring $S^0[\phi]$ out of the *rhs* of (5.4). Observing that the same ϕ kinetic action already appears on the *lhs*, we obtain ($\hbar = 1$)

$$S^{\rm int}_{\Lambda}[\phi] = -\log\left[\int_{C^\infty(M)_{(\Lambda,\Lambda_0]}} \mathcal{D}\chi\ \exp\left(-S^0[\chi] - S^{\rm int}_{\Lambda_0}[\phi,\chi]\right)\right] \tag{5.8}$$

which is the renormalization group equation for the effective *interactions*.

5.1.1 Running couplings and their β-functions

It should be clear that the partition function

$$\mathcal{Z}_\Lambda(g_i(\Lambda)) = \int_{C^\infty(M)_{\leq\Lambda}} \mathcal{D}\varphi\ \mathrm{e}^{-S^{\rm eff}_\Lambda[\varphi]/\hbar} \tag{5.9}$$

obtained from the effective action scale Λ (or at any lower scale) is exactly the same as the partition function we started with:

$$\mathcal{Z}_\Lambda(g_i(\Lambda)) = \mathcal{Z}_{\Lambda_0}(g_{i0};\Lambda_0) \tag{5.10}$$

because we're just performing the remaining integrals over the low–energy modes. In particular, as the scale is lowered infinitesimally (5.10) becomes the differential equation

$$\Lambda\frac{d\mathcal{Z}_\Lambda(g)}{d\Lambda} = \left(\Lambda\frac{\partial}{\partial\Lambda}\bigg|_{g_i} + \Lambda\frac{\partial g_i(\Lambda)}{\partial\Lambda}\,\frac{\partial}{\partial g_i}\bigg|_{\Lambda}\right)\mathcal{Z}_\Lambda(g) = 0\,. \tag{5.11}$$

Equation (5.11) is known as the renormalization group equation for the partition function, and is our first example of a *Callan–Symanzik equation.* It just says that as change the scale by integrating out modes, the couplings in the effective action $S^{\rm eff}_\Lambda$ vary to account for the change in the degrees of freedom over which we take the path integral, so that the partition function is in fact independent of the scale at which we define our theory, provided this scale is below our initial cut–off Λ_0.

The fact that the couplings themselves vary, or 'run', as we change the scale is an important notion. As we saw in zero and one dimensions, it's quite natural to expect the couplings to change as we integrate out modes, changing the degrees of freedom we can access at low scales. However, it seems strange: you've learned that the electromagnetic coupling

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137}\,.$$

What can it mean for the fine structure constant to depend on the scale? We'll understand the answer to such questions later.

With a generic initial action, the effective action will also take the general form

$$S_\Lambda^{\rm eff}[\phi] = \int {\rm d}^dx \, \left[\frac{Z_\Lambda}{2}\partial^\mu\phi\,\partial_\mu\phi + \sum_i \Lambda^{d-d_i} Z_\Lambda^{n_i/2}\, g_i(\Lambda)\,\mathcal{O}_i(x)\right] , \tag{5.12}$$

where the *wavefunction renormalization factor* Z_Λ accounts for the fact that it's perfectly possible for the coefficient of the kinetic term itself to receive quantum corrections as we integrate out modes. (Z_Λ is not to be confused with the partition function $\mathcal{Z}_\Lambda$!) At any given scale, we can of course define a *renormalized field*

$$\varphi := Z_\Lambda^{1/2}\phi \tag{5.13}$$

in terms of which the kinetic term will be canonically normalized. We've also included a power of $Z_\Lambda^{1/2}$ in the definition of our scale Λ couplings so that these powers are removed once one writes the action in terms of the renormalized field.

Since the running of couplings is so important, we give it a special name and define the *beta–function* β_i of the coupling g_i to be its derivative with respect to the logarithm of the scale:

$$\beta_i := \Lambda\frac{\partial g_i}{\partial\Lambda} \, . \tag{5.14}$$

The β-functions for dimensionless couplings take the form

$$\beta_i(g_j(\Lambda)) = (d_i - d)\, g_i(\Lambda) + \beta_i^{\rm quant}(g_j) \tag{5.15}$$

where the first term just compensates the variation of the explicit power of Λ in front of the coupling in (5.12). The second term $\beta_i^{\rm quant}$ represents the quantum effect of integrating out the high–energy modes. To actually compute this term requires us to perform the path integral and so will generically introduce dependence on all the other couplings in the original action (5.1), so that the β-function for g_i is a function of all the couplings $\beta_i(g_j)$.

Similarly, although at any given scale we can remove the wavefunction renormalization factor, moving to a different scale will generically cause it to re-emerge. We define the *anomalous dimension* γ_ϕ of the field ϕ by

$$\gamma_\phi := -\frac{1}{2}\Lambda\frac{\partial \ln Z_\Lambda}{\partial\Lambda} \tag{5.16}$$

Except for the fact that we're taken the derivative of the logarithm of $Z_\Lambda^{1/2}$, this is just the β-function for the coupling in front of the kinetic term. Like any β-function, γ_ϕ depends on the values of all the couplings in the theory. It gets it's name for reasons that will be apparent momentarily. If our theory contained more than one type of field, then we'd have a wavefunction renormalization factor and anomalous dimension for each field[21].

[21]In fact, in general we'd have a *matrix* of wavefunction renormalization factors, allowing different fields (of the same quantum numbers such as spin, charge *etc.*) to mix their identities as modes are integrated out.

5.1.2 Anomalous dimensions

Wavefunction renormalization plays an important role in correlation functions. Suppose we wish to compute the n–point correlator

$$\langle\phi(x_1)\cdots\phi(x_n)\rangle := \frac{1}{\mathcal{Z}}\int_{C^\infty(M)_{\leq\Lambda}} \mathcal{D}\phi\ \mathrm{e}^{-S_\Lambda^{\text{eff}}[Z_\Lambda^{1/2}\phi;\, g_i(\Lambda_0)]}\,\phi(x_1)\cdots\phi_n(x_n) \tag{5.17}$$

of fields inserted at points $x_1,\ldots,x_n \in M$ using the scale Λ theory, allowing for the possibility that we hadn't canonically normalized the field in the action. In terms of the canonically normalized field $\varphi := Z_\Lambda^{1/2}\phi$ this is

$$\langle\phi(x_1)\cdots\phi(x_n)\rangle = Z_\Lambda^{-n/2}\langle\varphi(x_1)\cdots\varphi(x_n)\rangle \tag{5.18}$$

since the change in the measure $\mathcal{D}\phi\to\mathcal{D}\varphi$ cancels as we've normalized by the partition function. Upon performing the φ path integral we will (in principle!) evaluate the remaining φ correlator as some function $\Gamma_\Lambda^{(n)}(x_1,\ldots,x_n;g_i(\Lambda))$ that depends on the scale Λ couplings and on the fixed points $\{x_i\}$.

Now, if the field insertions just involve modes with energies $\ll\Lambda$ then we should be able to compute the same correlator using just a lower scale theory — the operator insertions will be unaffected as we integrate out modes in the range $(s\Lambda,\Lambda]$ for some fraction $s<1$. Accounting for wavefunction renormalization gives

$$Z_{s\Lambda}^{-n/2}\,\Gamma_{s\Lambda}^{(n)}(x_1,\ldots,x_n;g_i(s\Lambda)) = Z_\Lambda^{-n/2}\,\Gamma_\Lambda^{(n)}(x_1,\ldots,x_n;g_i(\Lambda))\,, \tag{5.19}$$

or equivalently

$$\Lambda\frac{d}{d\Lambda}\Gamma_\Lambda^{(n)}(x_1,\ldots,x_n;g_i(\Lambda)) = \left(\Lambda\frac{\partial}{\partial\Lambda}+\beta_i\frac{\partial}{\partial g_i}+n\gamma_\phi\right)\Gamma_\Lambda^{(n)}(x_1,\ldots,x_n;g_i(\Lambda)) = 0 \tag{5.20}$$

infinitesimally. Equation (5.20) is the generalized Callan–Symanzik equation appropriate for correlation functions. Once again, it simply says that the couplings and wavefunction renormalization factors change as we lower the scale in such a way that correlation functions remain unaltered.

In a Poincaré invariant theory, correlation functions depend the distances between pairs of insertion points, as we saw in section (4.1.1). The typical size of these separations defines a new scale, quite apart from any choice of Λ, and we can use this to obtain an alternative interpretation of renormalization that is often useful. Integrate out modes in the range $(s\Lambda,\Lambda]$ as above, but having done so, let's now change coordinates on our space by $x^\mu\mapsto x'^\mu := sx$. The kinetic term $\int \mathrm{d}^dx\,(\partial\phi)^2$ is invariant under this scaling provided we take the field to transform as

$$\phi(sx) = s^{(2-d)/2}\phi(x)\,. \tag{5.21}$$

The remaining terms in the action are likewise unchanged by the rescaling provided we also rescale $\Lambda\to\Lambda/s$ in the opposite direction to x (as expected for an energy, rather than length, scale). Thus the energy scale $s\Lambda$ is restored to its original value Λ. It's important to realize that these scalings have nothing to do with integrating out degrees of freedom in the path integral; they're just scalings.

Under the combined operations we find

$$\begin{aligned}\Gamma^{(n)}_{\Lambda}(x_1,\ldots,x_n;g_i(\Lambda)) &= \left[\frac{Z_\Lambda}{Z_{s\Lambda}}\right]^{n/2}\,\Gamma^{(n)}_{s\Lambda}(x_1,\ldots,x_n;g_i(s\Lambda))\\ &= \left[s^{2-d}\frac{Z_\Lambda}{Z_{s\Lambda}}\right]^{n/2}\,\Gamma^{(n)}_{\Lambda}(sx_1,\ldots,sx_n;\,g_i(s\Lambda))\,,\end{aligned}\tag{5.22}$$

where the first line uses the result (5.19) of integrating out modes, while the second line shows how correlation functions are related under the rescaling. Notice that the couplings g_i and wavefunction renormalization in the final expression are evaluated at the point $s\Lambda$ appropriate for the low–energy theory: the numerical values of these couplings are not affected by our subsequent rescaling.

Equation (5.22) has an important interpretation. First, notice that if we'd *started* with insertions at points x_i/s then we could equivalently write

$$\Gamma^{(n)}_{\Lambda}(x_1/s,\ldots,x_n/s;g_i(\Lambda)) = \left[s^{2-d}\frac{Z_\Lambda}{Z_{s\Lambda}}\right]^{n/2}\,\Gamma^{(n)}_{\Lambda}(x_1,\ldots,x_n;\,g_i(s\Lambda))\,.\tag{5.23}$$

On the left stands a correlation function computed in the theory with couplings $g_i(\Lambda)$ where the separations between operators are $|x_i-x_j|/s$. Thus, as $s\to 0$ this correlator probes the long distance, or *infra–red* properties of the theory. We see from the *rhs* that such IR correlators may equivalently be obtained by studying a correlation function where all separations are held constant, but we compute using a theory with different values $g_i(s\Lambda)$ for the couplings. This makes perfect sense: the IR properties of the theory are governed by the low–energy modes that survive as we integrate out more and more high–energy degrees of freedom.

This equation also allows us to gain insight into the meaning of the anomalous dimension γ_ϕ. The power of $s^{n(2-d)/2}$ on the *rhs* of (5.22) is the classical scaling behaviour we'd expect for an object of mass dimension $n(d-2)/2$. Equation (5.22) shows that the net effect of integrating out high–energy modes is to modify the expected classical scaling by a simple factor depending on the wavefunction renormalization. To quantify this, set $s=1-\delta s$ with $0<\delta s\ll 1$. For each insertion of the field, (5.22) gives a factor

$$\left[s^{2-d}\frac{Z_\Lambda}{Z_{s\Lambda}}\right]^{1/2} = 1+\left[\frac{d-2}{2}+\gamma_\phi\right]\delta s+\cdots\tag{5.24}$$

with γ_ϕ as in (5.16). We see that the correlation function behaves *as if* the field scaled with mass dimension

$$\Delta_\phi=(d-2)/2+\gamma_\phi\tag{5.25}$$

rather than the classical value $(d-2)/2$. Δ_ϕ is known as the *scaling dimension* of the field ϕ, and the anomalous dimension γ_ϕ is the difference between this scaling dimension and the naive classical dimension.

5.2 Renormalization group flow

In this section we'll build up a general understanding of how theories change as we probe them in the infra-red. This conceptual understanding, first developed by Kadanoff and Wilson in the context of condensed matter field theory, will stand us in good stead when we come to renormalize theories perturbatively in later sections. Such calculations are often rather technical — the general picture of the present section will prevent us from getting bogged down in the details.

5.2.1 Renormalization group trajectories

To start to understand what happens under renormalization, let's suppose we start with a theory where all the β-functions vanish. That is, we consider a special action where the initial couplings are tuned to particular values $g_{i0} = g_i^*$ such that $\beta_j|_{g_j=g_i^*} = 0$, so that the couplings for this particular theory in fact do *not* depend on scale. Such a theory is known as a *critical point* of the RG flow. A simple example, called the *Gaussian critical point*, is just free theory where $g_i^* = 0$ for all terms in the action except for the (massless) kinetic term. Clearly, the β-functions all vanish at this Gaussian critical point, since the free theory has no interactions which could be responsible for generating vertices as the cut–off is lowered. However, by tuning the initial couplings very carefully, we may be able to cause non–trivial quantum corrections to cancel precisely the classical rescaling term in (5.15) so that the beta functions vanish. Thus it may be possible, though difficult, to find other critical points beyond the Gaussian one.

The couplings g_i^* being independent of scale has important implications for correlation functions. Firstly, note that since it is a dimensionless function of the other couplings, the anomalous dimension $\gamma_\phi(g_i^*) := \gamma_\phi^*$ is likewise scale independent at a critical point. Then for the two–point correlation function (5.20) becomes

$$\Lambda \frac{\partial \Gamma_\Lambda^{(2)}(x,y)}{\partial \Lambda} = -2\gamma_\phi^* \, \Gamma_\Lambda^{(2)}(x,y) \tag{5.26}$$

showing that $\Gamma^{(2)}$ is a homogeneous function of the scale. By Lorentz invariance it can depend on the insertion points only through $|x-y|$, and dimensional analysis shows that $\langle\phi(x)\phi(y)\rangle = \Lambda^{d-2}\, G(\Lambda|x-y|, g_i^*)$ for some function of the dimensionless combination $\Lambda|x-y|$ and the dimensionless couplings g_i^*. Thus, at a critical point the two–point function must take the form

$$\Gamma_\Lambda^{(2)}(x,y;\, g_i^*) = \frac{\Lambda^{d-2}}{\Lambda^{2\Delta_\phi}} \, \frac{c(g_i^*)}{|x-y|^{2\Delta_\phi}} \sim \frac{c(g_i^*)}{|x-y|^{2\Delta_\phi}} \tag{5.27}$$

in terms of the scaling dimension

$$\Delta_\phi = (d-2)/2 + \gamma_\phi \tag{5.28}$$

of ϕ, and where the constant $c(g_i^*)$ is independent of the insertion points. This power–law behaviour of correlation functions is characteristic of scale–invariant theories. In a theory where the interactions between the ϕ insertions was due to some massive state traveling from x to y, we'd expect the potential to decay as $\mathrm{e}^{-m|x-y|}/|x-y|$ where m is the mass of the intermediate state. As in electromagnetism, the pure power–law we have found for this correlator is a sign that our states are massless, so that their effects are long–range.

Critical theories are certainly very special. The metric appears in the action, so changing the metric leads to a change in the partition function given by

$$\delta g^{\mu\nu}(x)\,\frac{\delta}{\delta g^{\mu\nu}(x)}\ln\mathcal{Z} = -\left\langle\frac{\delta S}{\delta g^{\mu\nu}(x)}\right\rangle = -\delta g^{\mu\nu}(x)\,\langle T_{\mu\nu}(x)\rangle\,, \tag{5.29}$$

the expectation value of the stress tensor $T_{\mu\nu}$. If the metric transformation is just a scale transformation then $\delta g^{\mu\nu}\propto g^{\mu\nu}$, so scale invariance of a theory at a critical point g_i^* implies that $\langle T^{\mu}{}_{\mu}\rangle = 0$. In fact, all known examples of Lorentz–invariant, unitary QFTs that are scale invariant are actually invariant under the larger group of conformal transformations and it's believed that all critical points of RG flows are CFTs[22].

Now let's consider the behaviour of theories near to, but not at, a critical point. Since by definition the β-functions vanish when $g_i = g_i^*$, nearby we must have

$$\Lambda\frac{\partial g_i}{\partial\Lambda}\bigg|_{g_j^*+\delta g_j} = \mathrm{B}_{ij}\,\delta g_j + \mathcal{O}(\delta g^2) \tag{5.30}$$

where $\delta g_i = g_i - g_i^*$, and where B_{ij} is a constant (infinite dimensional!) matrix. Let σ_i be an eigenvector of B_{ij}, and let its eigenvalue be $\Delta_i - d$. Classically, we expected a dimensionless coupling to scale with a power of Λ determined by the explicit powers of Λ included in the action in (5.1), so that we'd have $\Delta_i = d_i$ classically. Just as for the correlation function in (5.20), the net effect of integrating out degrees of freedom is to modify this scaling so that near a critical point, the couplings really scale with a power of Λ determined by the eigenvalues of the linearized β-function matrix B_{ij}. The difference

$$\gamma_i := \Delta_i - d_i \tag{5.31}$$

is called the *anomalous dimension of the operator*, mimicking the anomalous dimension γ_ϕ of the field itself, while the quantity Δ_i itself is called the *scaling dimension* of the operator. If the quantum corrections vanished then the scaling dimension would coincide with the naive mass dimension of an operator obtained by counting the powers of fields and derivatives it contains.

Since σ_i is an eigenvector of B

$$\Lambda\frac{\partial\sigma_i}{\partial\Lambda} = (\Delta_i - d)\sigma_i + \mathcal{O}(\sigma^2) \tag{5.32}$$

and so the RG flow for σ_i is

$$\sigma_i(\Lambda) = \left(\frac{\Lambda}{\Lambda_0}\right)^{\Delta_i - d}\sigma_i(\Lambda_0) \tag{5.33}$$

at least to this order in the perturbation away from the critical point.

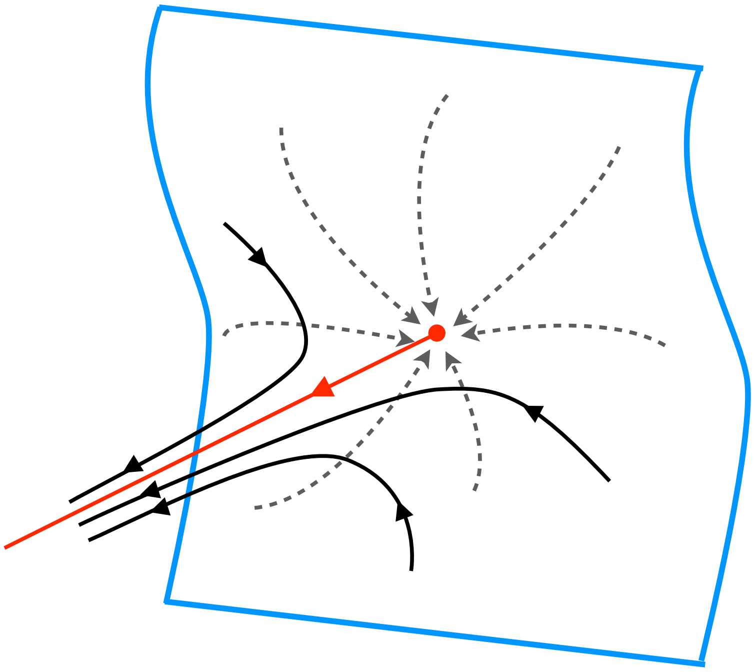

Figure 5: *Theories on the critical surface flow (dashed lines) to a critical point in the IR. Turning on relevant operators drives the theory away from the critical surface (solid lines), with flow lines focussing on the (red) trajectory emanating from the critical point.*

Now consider starting near a critical point and turning on the coupling to any operator with $\Delta_i > d$. According to (5.33) this coupling becomes smaller as the scale Λ is lowered, or as we probe the theory in the IR. We say that the corresponding operator is *irrelevant* since if we include it in the action then RG flow just makes us flow back to the critical point g_i^*. Classically, we can obtain operators with arbitrarily high mass dimension by including more and more fields and derivatives, so we expect that the critical point g_i^* sits on an infinite dimensional surface $\mathcal{C}$ such that if we turn on any combination of operators that move us along $\mathcal{C}$, under RG flow we will end up back at the critical point. $\mathcal{C}$ is known as the *critical surface* and we can think of the couplings of irrelevant operators as provided coordinates on $\mathcal{C}$, at least in the neighbourhood of g_i^*. (See figure 5.)

On the other hand, couplings with $\Delta_i < d$ grow as the scale is lowered and so are called *relevant*. If our action contains vertices with relevant couplings then RG flow will drive us away from the critical surface $\mathcal{C}$ as we head into the IR. Starting precisely from a critical point and turning on a relevant operator generates what is known as a *renormalized trajectory*: the RG flow emanating from the critical point. As we probe the theory at lower and lower scales we evolve along the renormalized trajectory either forever or until we eventually meet another[23] critical point g_i^{**}. Since each new field or derivative adds to the dimension of an operator, in fixed space–time dimension d there will be only finitely many

[22]It's a theorem that this is always true in two dimensions. It is believed to be true also in higher dimensions, but the question is actually a current hot topic of research.

[23]There are a few exotic examples where the theories flow to a limiting cycle rather than a fixed point.

(and typically only few) relevant operators, so the critical surface has finite codimension.

The remaining possibility is *marginal operators*, which have vanishing eigenvalues and so neither increase nor decrease under RG flow. At the Gaussian point, the scaling dimensions of operators are just given by their classical mass dimension, so we expect marginal operators to have scaling dimension $\Delta_i = d$. Near a critical point, quantum corrections can bring in a weak (typically logarithmic) dependence on scale to a classically marginal operator, making it either *marginally relevant* or *marginally irrelevant*. Provided the non–zero eigenvalues of these operators are sufficiently small, the size of such nearly marginal couplings can be unchanged for long periods of RG flow — although ultimately they will either be irrelevant or relevant. Such operators play an important role phenomenologically, as we will see.

A generic QFT will have an action that involves all types of operators and so lies somewhere off the critical surface. Under RG evolution, all the many irrelevant operators are quickly suppressed, while the relevant ones grow just as before. The flow lines of a generic theory thus strongly focus onto the renormalized trajectory, and so in the IR a generic QFT will closely resemble a theory emanating from the critical point, where only relevant operators have been turned on. The fact that many different high energy theories will flow to look the same in the IR is known as *universality*. It assures us that the properties of the theory in the IR are determined not by the infinite set of couplings $\{g_i\}$, but only by the couplings to a few relevant operators. We say that theories whose RG flows are all focussed onto the same trajectory emanating from a given critical point are in the same *universality class*. Theories in a given universality class could look very different microscopically, but will all end up looking the same at large distances. In particular, deep in the IR, these theories will all flow to the second critical point g_i^{**}. *This is the reason you can do physics!* To study a problem at a given energy scale you don't first need to worry about what the degrees of freedom at much higher energies are doing. They are, quite literally, irrelevant.

Let me emphasize that eigenvectors σ_i are generically *linear combinations* of the naive couplings in the action. Thus, turning on σ_i means we perturb away from the critical point by changing the couplings in front of the corresponding linear combination of our operators in the action. These RG 'eigenoperators' may be very different from any individual monomial in the fields you choose to include neatly in the effective interaction S^{eff}. A simple–looking individual operator $\mathcal{O}_i$ that appears in (5.1) or is explicitly inserted into a correlation function could actually consist of many RG eigenfunctions. We say that operators *mix*, because a given operator transforms under RG flow into its dominant eigenfunction, which could look very much more complicated.

5.2.2 Counterterms and the continuum limit

So far, we've considered a *fixed* initial theory $S_{\Lambda_0}[\phi]$ with initial couplings g_{i0}, and examined how these couplings change as we probe the theory at long distances. Our definition of the

Dimension	Relevant operators	Marginal operators
$d = 2$	ϕ^{2k} for all $k \geq 0$	$(\partial\phi)^2$, $\phi^{2k}(\partial\phi)^2$ for all $k \geq 0$
$d = 3$	ϕ^{2k} for $k = 1, 2$	$(\partial\phi)^2$, ϕ^6
$d = 4$	ϕ^2 for ≤ 3	$(\partial\phi)^2$, ϕ^4
$d > 4$	ϕ^2 for $0 \leq k \leq 3$	$(\partial\phi)^2$

Table 1: *Relevant & marginal operators in a Lorentz invariant theory of a single scalar field in various dimensions, near the Gaussian critical point where the classical dimensions of operators are a good guide. Only the operators invariant under $\phi \to -\phi$ are shown. Note that the kinetic term $(\partial\phi)^2$ is always marginal, and the mass term ϕ^2 is always relevant.*

low–energy effective action as

$$S_\Lambda^{\text{eff}}[\phi] := -\hbar \log \left[\int_{C^\infty(M)_{(\Lambda,\Lambda_0]}} \mathcal{D}\chi \ \exp\left(-S_{\Lambda_0}[\phi + \chi]/\hbar\right) \right] \tag{5.34}$$

ensured that the partition function and correlation functions of low–energy observables were independent of the scale Λ. The question remains: what about dependence on the initial cut–off Λ_0? In this section we'll examine this by asking a sort of converse: Suppose we fix a particular low–energy theory (perhaps motivated by the results of some finite–scale experiments). How can we remove the high–energy cut–off, sending $\Lambda_0 \to \infty$, without affecting what the theory predicts for low–energy phenomena. We call this *taking the continuum limit* of our theory, since sending $\Lambda_0 \to \infty$ is allowing the field to fluctuate on arbitrarily small scales.

The key to achieving this lies with the universality of the renormalization group flow. First, suppose our initial couplings g_{i0} happen to lie on the critical surface $\mathcal{C}$, within the domain of attraction of g_i^*. Then as we raise the cut–off Λ_0, the theory we obtain at any fixed scale Λ will be driven to the critical point g_i^* as all the irrelevant operators become arbitrarily suppressed by positive powers of Λ/Λ_0. The critical point is a fixed point of the RG flow and is scale invariant, so we can happily send $\Lambda_0 \to \infty$. More precisely, whenever the theory S_{Λ_0} lives on the critical surface, the limit

$$\lim_{\Lambda_0 \to \infty} \left[\int_{C^\infty(M)_{(\Lambda,\Lambda_0]}} \mathcal{D}\chi \ \exp\left(-S_{\Lambda_0}[\phi + \chi]/\hbar\right) \right] , \tag{5.35}$$

exists, provided we take this limit *after* computing the path integral. The resulting scale-Λ effective theory will be a CFT, independent of Λ. Since $\mathcal{C}$ has only finite codimension, we only have to tune finitely many coefficients (those of all the relevant operators) in order to ensure that $g_{i0} \in \mathcal{C}$.

Theories such as Yang–Mills or QCD are *not* CFTs, but rather have relevant (and marginally relevant) operators turned on in their actions. How then can we understand the continuum limit of such theories? Consider a theory whose initial conditions are near, but not on $\mathcal{C}$. Universality of the RG flow shows that as we head into the IR, such a theory flows towards the critical point g_i^* for a while, but eventually diverges away, focussing on a renormalized trajectory as in figure 5. Let μ denote the energy scale at which this theory passes closest to g_i^*. Since RG flow is determined by the initial conditions, μ depends only on the theory we started with. On dimensional grounds we must have

$$\mu = \Lambda_0 \, f(g_{i0}) \tag{5.36}$$

where $f(g_{i0})$ is some function of the dimensionless couplings $\{g_{i0}\}$ and $f = 0$ on $\mathcal{C}$, since all these theories flow to g_i^* exactly. To obtain a theory with relevant or marginal operators, we tune the initial couplings $\{g_{i0}\}$ so that μ remains *finite* as we take $\Lambda_0 \to \infty$. If $\operatorname{codim}(\mathcal{C}) = r$ then this is one condition on r parameters — the coefficients of the relevant operators in the initial action. The theory we end up with thus depends on $(r-1)$ parameters, together with the scale μ.

We achieve this tuning by introducing new *counterterms* $S_{\mathrm{CT}}[\varphi, \Lambda_0]$ that depend on the fields ϕ as well as explicitly on the cut–off Λ_0, modifying the initial action to

$$S_{\Lambda_0}[\varphi] \to S_{\Lambda_0}[\varphi] + S_{\mathrm{CT}}[\varphi, \Lambda_0] \,. \tag{5.37}$$

The effective actions we considered before already contained all possible monomials in fields and their derivatives, so in this sense the counterterms add nothing new. However, the values of the counterterm couplings are to be chosen by hand — varying these couplings thus *changes* which initial high–energy theory we're considering, as opposed to running a set of couplings under RG flow, which just describes how the *same* theory appears at different scales. The counterterms are tuned so that the limit

$$\mathrm{e}^{-S_\Lambda^{\mathrm{eff}}[\phi]/\hbar} = \lim_{\Lambda_0 \to \infty} \left[\int_{C^\infty(M)_{(\Lambda,\Lambda_0]}} \mathcal{D}\chi \; \exp\left(-\frac{S[\phi+\chi]}{\hbar} - S_{\mathrm{CT}}[\phi+\chi, \Lambda_0] \right) \right] \tag{5.38}$$

exists. Notice again that the limit is taken *after* performing the path integral. Sending $\Lambda_0 \to \infty$ defines a continuum QFT with finite (or *renormalized*) relevant couplings at scale Λ.

The reason for making S_{CT} explicit, rather than just treating the counterterms as a modification $\{g_{i0}\}$, is that in practice we work perturbatively. To evaluate the path integral in (5.38), we first compute quantum corrections to 1-loop order using the original action S. These 1-loop corrections will depend on the cut–off Λ_0, and will be proportional to $\hbar$. In general, they will diverge as $\Lambda_0 \to \infty$ reflecting the fact that we lose control of the original theory if the cut–off is removed naïvely. However, vertices in S_{CT} provide further contributions to these quantum corrections. By tuning the values of the couplings in S_{CT} by hand, we can obtain a finite limit. Notice that S_{CT} comes with one extra power of $\hbar$ in (5.38) compared to the original action. Thus, quantum corrections to the effective action arising from 1-loop diagrams of the original action should be matched by the *tree–level* contributions from S_{CT}. We'll get plenty of practice in doing this in the following sections.

There's one further possibility to consider. Suppose that to explain some experimental result, be it the scattering of pions and nucleons or the falling of apples, we need our low–energy theory to contain irrelevant operators. If we really try to take the cut–off $\Lambda_0 \to \infty$, such operators will be arbitrarily suppressed at any finite energy scale. So their presence indicates that our theory cannot be valid up to arbitrarily high energies; there must be a finite energy scale at which new physics comes in to play. In the case of pion–nucleon scattering, this scale is ~ 217 MeV and indicates the presence of quarks, gluons and the whole structure of QCD. For radioactive β-decay, the scale is ~ 250 GeV and indicated the electroweak theory, while for gravity the scale is $\sim 10^{19}$ GeV, where probably the whole notion of QFT itself must give way. Perhaps most interesting of all are *marginally* irrelevant operators, like the quartic coupling (Φ^4) of the Higgs in the Standard Model. Strictly speaking, just like irrelevant operators, marginally irrelevant operators are arbitrarily suppressed as the cut–off is removed. However, they typically decay only logarithmically as Λ_0 is raised, rather than as a power law. Such operators thus afford us a tiny glimpse of new physics at exponentially high energy scales, far beyond the range of current accelerators.

5.3 Calculating RG evolution

It's time to think about how to calculate the quantum corrections to β-functions generated as we integrate out high energy modes. I'll this section I want to explain this in a way that I think is conceptually clear, and the natural generalization of what we have already seen in zero and one dimension. However, I'll warn you in advance that the techniques here are not the most convenient way to calculate β-functions.

5.3.1 Polchinski's equation

In perturbation theory, the *rhs* of (5.8) may be expanded as an infinite series of connected Feynman diagrams. If we wish to compute the low–energy effective interaction $S_\Lambda^{\text{int}}[\phi]$ as an integral over space–time in the usual way, then we should use the position space Feynman rules. As in section 3.4, the position space propagator $D^{(\chi)}(x,y)$ for the high–energy field χ is

$$D^{(\chi)}(x,y) = \int_{\Lambda<|p|\leq\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d}\, \frac{\mathrm{e}^{\mathrm{i}p\cdot(x-y)}}{p^2+m^2} \tag{5.39}$$

where we note the restriction to momenta in the range $\Lambda < |p| \leq \Lambda_0$. As usual, vertices from the high–energy action $S_{\Lambda_0}^{\text{int}}[\phi,\chi]$ come with an integration $\int \mathrm{d}^d x$ over their location that imposes momentum conservation at the vertex. Now, diagrams that exclusively involve vertices which are independent of ϕ contribute just to a field–independent term on the *lhs* of (5.8). This term represents the shift in vacuum energy due to integrating out the χ field; we will henceforth ignore it[24]. The remaining diagrams use vertices including at least one ϕ field, treated as external. Evaluating such a diagram leads to a contribution to the effective interaction $S_\Lambda^{\text{int}}[\phi]$ at scale Λ.

For general scales Λ and Λ_0 equation (5.8) is extremely difficult to handle; the integral on the right is a full path integral in an interacting theory. To make progress we consider

[24]This is harmless in a non–gravitational theory, but is really the start of the **cosmological constant problem**.

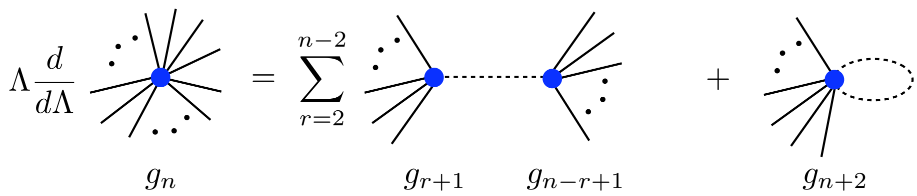

Figure 6: *A schematic representation of the renormalization group equation for the effective interactions when the scale is lowered infinitesimally. Here the dashed line is a propagator of the mode with energy Λ that is being integrated out, while the external lines represent the number of low–energy fields at each vertex. All these external fields are evaluated at the same point x The total number of fields attached to a vertex is indicated by the subscripts on the couplings g_i.*

the case that we lower the scale only infinitesimally, setting $\Lambda = \Lambda_0 - \delta\Lambda$. To lowest order in $\delta\Lambda$, the χ propagator reduces to[25]

$$D^{(\chi)}_{\Lambda}(x,y) = \frac{1}{(2\pi)^d}\,\frac{\Lambda^{d-1}\,\delta\Lambda}{\Lambda^2+m^2}\int_{S^{d-1}} \mathrm{d}\Omega\ \mathrm{e}^{\mathrm{i}\Lambda\hat{p}\cdot(x-y)} \tag{5.40}$$

as the range of momenta shrinks down, where $\mathrm{d}\Omega$ denotes an integral over a unit $(d-1)$-sphere in momentum space. This is a huge simplification! Since every χ propagator comes with a factor of $\delta\Lambda$, to lowest order in $\delta\Lambda$ *we need only consider diagrams with a single χ propagator*. Since ϕ is treated as an external field, we have only two possible classes of diagram: either the χ propagator links together two separate vertices in $S^{\mathrm{int}}[\phi,\chi]$ or else it joins a single vertex to itself.

This diagrammatic represention of the process of integrating out degrees of freedom is shown in figure 6. It has a very clear intuitive meaning. The mode χ appearing in the propagator is the highest energy mode in the original scale Λ_0 theory. It thus probes the shortest distances we can reliably access using $S^{\mathrm{eff}}_{\Lambda_0}$. When we integrate this mode out, we can no longer resolve distances $1/\Lambda_0$ and our view of the 'local' interaction vertices is correspondingly blurred. The graphs on the *rhs* of figure 6 represent new contributions to the n–point ϕ vertex in the lower scale theory coming respectively from two nearby vertices joined by a χ field, or a higher point vertex with a χ loop attached. Below scale Λ_0 we image that we are unable to resolve the short distance χ propagator.

We can write an equation for the change in the effective action that captures the information in the Feynman diagrams in figure 6. It was obtained by Polchinski[26], and is really just the infinitesimal version of Wilson's renormalization group equation (5.4) for the effective action. Polchinski's equation is

$$-\Lambda\frac{\partial S^{\mathrm{int}}_{\Lambda}[\phi]}{\partial\Lambda} = \int \mathrm{d}^d x\,\mathrm{d}^d y\left[\frac{\delta S^{\mathrm{int}}_{\Lambda}}{\delta\phi(x)}\,D_{\Lambda}(x,y)\,\frac{\delta S^{\mathrm{int}}_{\Lambda}}{\delta\phi(y)} - D_{\Lambda}(x,y)\frac{\delta^2 S^{\mathrm{int}}_{\Lambda}}{\delta\phi(x)\,\delta\phi(y)}\right], \tag{5.41}$$

[25] To lowest order, it doesn't matter whether we use Λ_0 or Λ in this expression.

[26] Polchinski actually wrote a slightly more general version of the momentum space version of this equation, including source terms $\int J\phi$ in the effective action.

where $D_\Lambda(x,y)$ is the propagator (5.40) for the mode at energy Λ that is being integrated out. The variations of the effective interactions tell us how this propagator joins up the various vertices. Notice in the second term that since $S^{\rm int}[\phi]$ is local, both the $\delta/\delta\phi$ variations must act at the same place if we are to get a non–zero result. On the other hand, the first term generates non–local contributions to the effective action since it links fields at x to fields at a different point y. In position space we expect a propagator at scale Λ^2+m^2 to lead to a potential $\sim \mathrm{e}^{-\sqrt{\Lambda^2+m^2}\,r}/r^{d-3}$ so this non–locality is mild and we can expanding the fields in $\delta S^{\rm int}/\delta\phi(y)$ as a series in $(x-y)$. This leads to new contributions to interactions involving derivatives of the fields, just as we saw in section 3.3 in one dimension. Finally, the minus signs in (5.41) comes from expanding $\mathrm{e}^{-S^{\rm int}[\phi]}$ to obtain the Feynman diagrams.

It's convenient to rewrite Polchinski's equation (5.41) as

$$\frac{\partial}{\partial t}\mathrm{e}^{-S^{\rm int}[\phi]} = -\int \mathrm{d}^dx\,\mathrm{d}^dy\; D_\Lambda(x,y)\,\frac{\delta^2}{\delta\phi(x)\,\delta\phi(y)}\mathrm{e}^{-S^{\rm int}[\phi]}\,, \tag{5.42}$$

in which form it reveals itself as a form of heat equation, with renormalization group 'time'[27] $t \equiv \ln\Lambda$ and 'Laplacian'

$$\Delta = \int \mathrm{d}^dx\,\mathrm{d}^dy\; D_\Lambda(x,y)\,\frac{\delta^2}{\delta\phi(x)\,\delta\phi(y)} \tag{5.43}$$

on the space of fields. Heat flow on a Riemannian manifold N is a strongly *smoothing* operation: if we expand a function $f : N\times\mathbb{R}_{>0}\to\mathbb{R}$ as

$$f(x,t) = \sum_k \tilde{f}_k(t)\,u_k(x)$$

in terms of a basis of eigenfunctions $u_k(x)$ of the Laplacian on N, then under heat flow the coefficients evolve as $\tilde{f}_k(t) = \tilde{f}_k(0)\,\mathrm{e}^{-\lambda_k t}$. Consequently, all components $\tilde{f}_k(t)$ corresponding to positive eigenvalues λ_k are quickly damped away, with only the constant piece (with zero eigenvalue) surviving. This just corresponds to the well–known fact that a heat spreads out from areas of high concentration (such as a flame) until the whole room is at constant temperature. On a manifold with a pseudo–Riemannian (rather than Riemannian) manifold, some eigenvalues can be negative. These functions would then be enhanced under heat flow. Exactly the same thing happens under RG flow. Eigenfunctions of the Laplacian in (5.42) are combinations of operators in the effective interactions. Depending on the sign of their corresponding eigenvalues, these operators will be either enhanced or suppressed under the flow.

5.3.2 The local potential approximation

Polchinski's equation contains exact information about the behaviour of every possible operator under RG flow. Unfortunately, while it's structurally simple, actually solving this equation as it stands is prohibitively difficult, so we seek a more managable approximation.

[27] In the AdS/CFT correspondence, this RG time really does turn into an honest direction: into the bulk of anti–de Sitter space!

To obtain one, observe that except for the kinetic term, operators involving derivatives are irrelevant[28] whenever $d > 2$. This suggests that we can restrict attention to actions of the form

$$S_\Lambda^{\rm eff}[\varphi] = \int \mathrm{d}^d x \, \left[\frac{1}{2}\partial^\mu\varphi\,\partial_\mu\varphi + V(\varphi)\right] \tag{5.44}$$

where the potential

$$V(\varphi) = \sum_k \Lambda^{d-k(d-2)}\frac{g_{2k}}{(2k)!}\,\varphi^{2k} \tag{5.45}$$

does not involve derivatives of ϕ. For simplicity, we've chosen $V(-\varphi) = V(\varphi)$, while the couplings g_{2k} are dimensionless as before. Neglecting the derivative interactions is known as the *local potential approximation*; it's important because it will tell us the shape of the effective potential experienced by a slowly varying field. Splitting the field $\varphi = \phi + \chi$ into its low– and high–energy modes as before, we expand the action as an infinite series

$$S_\Lambda^{\rm eff}[\phi+\chi] = S_\Lambda^{\rm eff}[\phi] + \int \mathrm{d}^d x\,\left[\frac{1}{2}(\partial\chi)^2 + \frac{1}{2}\chi^2\,V''(\phi) + \frac{1}{3!}\chi^3\,V'''(\phi) + \cdots\right]. \tag{5.46}$$

Notice that we have chosen a definition of ϕ so that it sits at a minimum of the potential, $V'(\phi) = 0$. This can always be arranged by adding a constant to ϕ, which is certainly a low–energy mode.

Now consider integrating out the high–energy modes χ. As before, we lower the scale infinitesimally, setting $\Lambda' = \Lambda - \delta\Lambda$ and working just to first order in $\delta\Lambda$. In any given Feynman graph, each χ loop comes with an integral of the form

$$\int_{\Lambda-\delta\Lambda<|p|\leq\Lambda}\frac{\mathrm{d}^d p}{(2\pi)^d}\,(\cdots) = \delta\Lambda\,\frac{\Lambda^{d-1}}{(2\pi)^d}\int_{S^{d-1}}\mathrm{d}\Omega\,(\cdots)$$

where $\mathrm{d}\Omega$ denotes an integral over a unit $S^{d-1}\subset\mathbb{R}^d$ and $(\cdots)$ represents the propagators and vertex factors involved in this graph. As with Polchinski's equation, since each loop integral comes with a factor of $\delta\Lambda$, to lowest non–trivial order in $\delta\Lambda$ we need consider at most 1-loop diagrams for χ.

Suppose a particular graph involves an number v_i vertices containing i powers of χ and arbitrary powers of ϕ. Euler's identity tells us that a connected graph with e edges and ℓ loops obeys

$$e - \sum_i v_i = \ell - 1\,, \tag{5.47}$$

Since we're only integrating over the high scales modes, χ is the only propagating field. Furthermore, since we're integrating out χ completely, there are no external χ lines. Thus we also have the identity

$$2e = \sum_i i\,v_i \tag{5.48}$$

since every χ propagator is emitted and absorbed at some (not necessarily distinct) vertex. Eliminating e from (5.47) gives

$$\ell = 1 + \sum_i \frac{i-2}{2}\,v_i\,. \tag{5.49}$$

[28] At least near the Gaussian critical point where classical scaling dimensions are a reasonable guide.

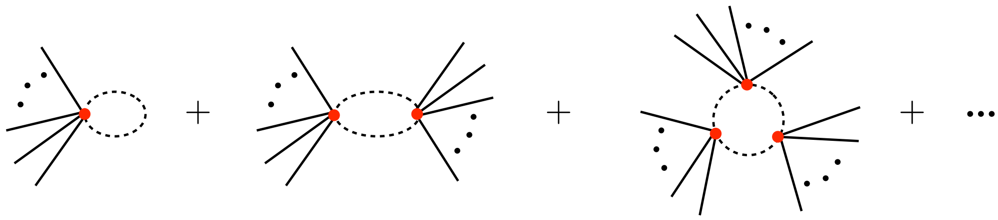

Figure 7: *Diagrams contributing in the local potential approximation to RG flow. The dashed line represents a χ propagator with $|p| = \Lambda$, while the solid lines represent external ϕ fields. All vertices are quadratic in χ.*

We only want to keep track of 1-loop diagrams, so we see that only the vertices with $i = 2$ χ lines (and arbitrary numbers of ϕ lines) are important. We can thus truncate the difference $S^{\text{eff}}_{\Lambda}[\phi+\chi] - S^{\text{eff}}_{\Lambda}[\phi]$ in (5.46) to

$$S^{(2)}[\chi] = \int \mathrm{d}^d x \left[\frac{1}{2}\partial\chi^2 + \frac{1}{2}V''(\phi)\,\chi^2\right] \tag{5.50}$$

so that χ appears only quadratically.

The diagrams that can be constructed from this action are shown in figure 7. If we make the temporary assumption that the low–energy field ϕ is actually constant, then in momentum space the quadratic action $S^{(2)}$ becomes

$$\begin{aligned} S^{(2)}[\chi] &= \int_{\Lambda-\delta\Lambda<|p|\leq\Lambda} \frac{\mathrm{d}^d p}{(2\pi)^d}\,\tilde{\chi}(p)\left[\frac{1}{2}p^2 + \frac{1}{2}V''(\phi)\right]\tilde{\chi}(-p) \\ &= \frac{\Lambda^{d-1}\delta\Lambda}{2(2\pi)^d}\left(\Lambda^2 + V''(\phi)\right)\int_{S^{d-1}} \mathrm{d}\Omega\;\tilde{\chi}(\Lambda\hat{p})\,\tilde{\chi}(-\Lambda\hat{p}) \end{aligned} \tag{5.51}$$

using the fact that these modes have energies in a narrow shell of width $\delta\Lambda$.

Performing the path integral over χ is now straightforward. If the narrow shell contains N momentum modes, then from standard Gaussian integration

$$\mathrm{e}^{-\delta_\Lambda S^{\text{eff}}[\phi]} = \int \mathcal{D}\chi\,\mathrm{e}^{-S_2[\chi,\phi]} = C\left(\frac{\pi}{\Lambda^2 + V''(\phi)}\right)^{N/2}. \tag{5.52}$$

On a non–compact manifold, N is actually infinite. To regularize it, we place our theory in a box of linear size L and impose periodic boundary conditions. The momentum is then quantized as $p_\mu = 2\pi n_\mu / L$ for $n_\mu \in \mathbb{Z}$ so that there is one mode per $(2\pi)^d$ volume in Euclidean space–time. The volume of space–time itself is L^d. Thus

$$N = \frac{\mathrm{Vol}(S^{d-1})}{(2\pi)^d}\,\Lambda^{d-1}\delta\Lambda\,L^d \tag{5.53}$$

which diverges as the volume L^d of space–time becomes infinite. However, we can obtain a (correct) finite answer once we recognize that the cause of this divergence was ou simplifying assumption that ϕ was constant. For spatially varying ϕ, we would instead obtain

$$\delta_\Lambda S^{\rm eff}[\phi] = a\,\Lambda^{d-1}\delta\Lambda \int \mathrm{d}^d x\,\ln\left[\Lambda^2 + V''(\phi)\right] \tag{5.54}$$

where the factor of $L^d \times \ln[\Lambda^2 + V''(\phi)]$ in (5.52) has been replaced by an integral over M. The constant

$$a := \frac{\mathrm{Vol}(S^{d-1})}{2(2\pi)^d} = \frac{1}{(4\pi)^{d/2}\,\Gamma(d/2)} \tag{5.55}$$

is proportional to the surface area of a $(d-1)$-dimensional unit sphere. Expanding the *rhs* of (5.54) in powers of ϕ leads to a further infinite series of ϕ vertices which combine with those present at the classical level in $V(\phi)$. Once again, integrating out the high–energy field χ has lead to a modification of the couplings in this potential.

We're now in position to write down the β-functions. Including the contribution from both the classical action and the quantum correction (5.54), the β-function for the ϕ^{2k} coupling is

$$\Lambda\frac{dg_{2k}}{d\Lambda} = [k(d-2)-d]g_{2k} - a\Lambda^{k(d-2)}\,\frac{\partial^{2k}}{\partial\phi^{2k}}\ln\left[\Lambda^2 + V''(\phi)\right]\bigg|_{\phi=0}\,. \tag{5.56}$$

For instance, the first few terms in this expansion give

$$\begin{aligned}
\Lambda\frac{dg_2}{d\Lambda} &= -2g_2 - \frac{ag_4}{1+g_2}\\
\Lambda\frac{dg_4}{d\Lambda} &= (d-4)g_4 - \frac{ag_6}{1+g_2} + \frac{3ag_4^2}{(1+g_2)^2}\\
\Lambda\frac{dg_8}{d\Lambda} &= (2d-6)g_6 - \frac{ag_8}{1+g_2} + \frac{15ag_4g_6}{(1+g_2)^2} - \frac{30ag_4^3}{(1+g_2)^3}
\end{aligned} \tag{5.57}$$

as β–functions for the mass term, ϕ^4 and ϕ^6 vertices.

There are several things worth noticing about the expressions in (5.57). Firstly, each term on the right comes from a particular class of Feynman graph; the first term is the scaling behaviour of the classical ϕ^{2k} vertex, the second term involves a single χ propagator with both ends joined to the same valence $2k+2$ vertex, the third (when present) involves a pair of χ propagators joining two vertices of total valence $2k+4$, *etc.*. Secondly, we note that these Feynman diagrams are different to the ones that appeared in (5.41). By taking the local potential approximation, we have neglected any possible derivative terms that may have contributed to the running of the couplings in $V(\phi)$. The effect of this is seen in the higher–order terms that appear on the *rhs* of (5.57). From the point of view of the Wilson–Polchinski renormalization group equation, the local potential approximation effectively amounts to *solving* the β-function equations that follow from (5.41), writing the derivative couplings in terms of the non–derivative ones, and then substituting these back into the remaining β-functions for non–derivative couplings to obtain (5.57). The message is that the price to be paid for ignoring possible couplings in the effective action is more complicated β-functions. We will see this again in chapter **??**, where β-functions will no longer be determined purely at one loop.

Finally, recall that $g_2 = m^2/\Lambda^2$ is the mass of the ϕ field in units of the cut–off. If this mass is very large, so $g_2 \gg 1$, then the quantum corrections to the β-functions in (5.57) are strongly suppressed. As for correlation functions near to, but not at, a critical point, this is as we would expect. A particle of mass m leads to a potential $V(r) \sim \mathrm{e}^{-mr}/r^{d-3}$ in position space, so should not affect physics on scales $r \gg m^{-1}$.

5.3.3 The Gaussian critical point

From our discussion in section 5.2, we expect that the limiting values of the couplings in the deep IR will be a critical point of the RG evolution (5.56). The simplest type of critical point is the Gaussian fixed–point where $g_{2k} = 0 \; \forall \; k > 1$, corresponding to a free theory. Every one of the Feynman diagrams shown on the right of the Wilson renormalization group equation in figure 6 involves a vertex containing at least three fields (either χ or ϕ), so if we start from a theory where the couplings to each of these vertices are precisely set to zero, then no interactions can ever be generated. Indeed, (5.57) shows that the Gaussian point is indeed a fixed–point of the RG flow, with the mass term β-function $\beta_2 = -2g_2$ simply compensating for the scaling of the explicit power of Λ introduced to make the coupling dimensionless.

Last term you used perturbation theory to study ϕ^4 theory in four dimensions. Using perturbation theory means that you considered this theory in the neighbourhood of the Gaussian critical point so that the couplings could be treated as 'small'. Let's examine this again using our improved understanding of RG flow. Firstly, to find the behaviour of any coupling near to the free theory, as in equation (5.30) we should linearize the β-functions around the critical point. We'll use our results (5.57) for a theory with an arbitrary polynomial potential $V(\phi)$. To linear order in the couplings, only the first two terms on the *rhs* of (5.57) contribute, giving

$$\beta_{2k} = \Lambda \frac{\partial g_{2k}}{\partial \Lambda} = (k(d-2) - d)\, g_{2k} - a g_{2k+2} \tag{5.58}$$

where $\delta g_{2k} = g_{2k} - g^*_{2k} = g_{2k}$ since $g^*_{2k} = 0$ for the Gaussian critical point. Writing the linearized β-functions in the form $\beta_i = \mathrm{B}_{ij} g_j$ we see that the matrix B_{ij} is upper triangular, so its eigenvalues are simply its diagonal entries $k(d-2) - d$. In four dimensions, these eigenvalues are $2k - 4$, which is positive for $k \geq 3$. Thus, in four dimensions, deforming the free action by an operator of the form ϕ^{2k} with $k \geq 3$ is an *irrelevant* perturbation: turning on any such operator takes us away from the free theory in a direction along the critical surface, and we are pushed back to the free theory as the cut–off is lowered.

On the other hand, the mass term g_2 is a *relevant* deformation of the free action. Turning on even arbitrarily small masses will lead us away from the massless theory as the cut–off is lowered. Of course, once g_2 is large we cannot trust our linearized approximation (5.58), and the correct result (5.57) shows that the quantum corrections to g_2 are eventually suppressed as the mass becomes large in units of the cut–off.

The remaining coupling g_4 is particularly interesting. We've seen that for $k \geq 3$, the ϕ^{2k} interactions are irrelevant in $d = 4$ near the Gaussian fixed point, so at low energies we may neglect them. β_4 then vanishes to linear order, so that the ϕ^4 coupling is marginal at this order. To study its behaviour, we need to go to higher order. From (5.57) we have

$$\beta_4 = \Lambda \frac{\partial g_4}{\partial \Lambda} = \frac{3}{16\pi^2}\, g_4^2 + \mathcal{O}(g_4^2 g_2) \tag{5.59}$$

to quadratic order, where we've again dropped the g_6 term. Equation (5.59) is solved by

$$\frac{1}{g_4(\Lambda)} = C - \frac{3}{16\pi^2} \ln \Lambda \tag{5.60a}$$

where C is an integration constant. Equivalently, we may write

$$g_4(\Lambda) = \frac{16\pi^2}{3\ln(\mu/\Lambda)} \tag{5.60b}$$

in terms of some arbitrary scale μ. Since g_4 is the coefficient of the highest power of ϕ that appears in our potential, we must have $g_4 > 0$ if the action is to be bounded as $|\phi| \to \infty$, so we must choose $\mu > \Lambda$.

There are several important things to learn from this result. Firstly, we see that $g_4(\Lambda)$ decreases as $\Lambda \to 0$, ultimately being driven to zero. However, the scale dependence of g_4 is rather mild; instead of power–law behaviour we have only logarithmic dependence on the cut–off. Thus the ϕ^4 coupling, which was marginal at the classical level, because marginally irrelevant once quantum effects are taken into account. In the deep IR, we see only a free theory.

Secondly, away from the IR we notice that the integration constant μ determines a scale at which the coupling diverges. If we try to follow the RG trajectories back into the UV, perturbation theory will certainly break down before we reach $\Lambda \approx \mu$. The fact that the couplings in the action can be traded for energy scales μ at which perturbation theory breaks down is a ubiquitous phenomenon in QFT known as **dimensional transmutation**. We'll meet it many times in later chapters. The question of whether the ϕ^4 coupling *really* diverges as we head into the UV or just *appears to* in perturbation theory is rather subtle. More sophisticated treatments back up the belief that it does indeed diverge: in the UV we lose all control of the theory and in fact we do not believe that ϕ^4 theory really exists as a well–defined continuum QFT in four dimensions. This has important phenomenological implications for the Standard Model, through the quartic coupling of the scalar Higgs boson; take the Part III Standard Model course if you want to find out more.

The fact that the ϕ^4 coupling is not a free constant, but is determined by the scale and can even diverge at a finite scale $\Lambda = \mu$ should be worrying. How can we ever trust perturbation theory? The final lesson of (5.60b) is that if we want to use perturbation theory, we should always try to choose our cut–off scale so as to make the couplings as small as possible. In the case of ϕ^4 theory this means we should choose Λ as low as possible. In particular, if we want to study physics at a particular length scale ℓ, then our best chance for a weakly coupled description is to integrate out all degrees of freedom on length scales shorter than ℓ, so that $\Lambda \sim \ell^{-1}$.

5.3.4 The Wilson–Fisher critical point

The conclusion at the end of the previous section was that ϕ^4 theory does not have a continuum limit in $d=4$. Since the only critical point is the Gaussian free theory we reach at low energies, four dimensional scalar theory is known as a **trivial** theory.

It's interesting to ask whether there are other, non–trivial critical points away from four dimensions. In general, finding non–trivial critical points is a difficult problem. Wilson and Fisher had the idea of introducing a parameter $\varepsilon := 4-d$ which is treated as 'small' so that one is 'near' four dimensions. One then hopes that results obtained via the **ε–expansion** may remain valid in the physically interesting cases of $d=3$ or even $d=2$. From the local potential approximation (5.56) Wilson & Fisher showed that there is a critical point $g_i^{\rm WF}$ where

$$g_2^{\rm WF} = -\frac{1}{6}\varepsilon + \mathcal{O}(\varepsilon^2)\,, \qquad\qquad g_4^{\rm WF} = \frac{1}{3a}\varepsilon + \mathcal{O}(\varepsilon^2) \tag{5.61}$$

and $g_{2k}^{\rm WF} \sim \varepsilon^k$ for all $k>2$. We require $\varepsilon > 0$ to ensure that $V(\phi)\to 0$ as $|\phi|\to\infty$ so that the theory can be stable.

To find the behaviour of operators near to this critical point, once again we linearize the β-functions of (5.57) around $g_{2k}^{\rm WF}$. Truncating to the subspace spanned by (g_2, g_4) we have

$$\Lambda\frac{\partial}{\partial\Lambda}\begin{pmatrix}\delta g_2\\ \delta g_4\end{pmatrix} = \begin{pmatrix}\varepsilon/3-2 & -a(1+\varepsilon/6)\\ 0 & \varepsilon\end{pmatrix}\begin{pmatrix}\delta g_2\\ \delta g_4\end{pmatrix}\,. \tag{5.62}$$

The matrix has eigenvalues $\varepsilon/3-2$ and ε, with corresponding eigenvectors

$$\sigma_2 = \begin{pmatrix}1\\0\end{pmatrix}\,, \qquad\qquad \sigma_4 = \begin{pmatrix}-a(3+\varepsilon/2)\\ 2(3+\varepsilon)\end{pmatrix} \tag{5.63}$$

respectively. In $d=4-\varepsilon$ dimensions we have

$$a = \left.\frac{1}{(4\pi)^{d/2}}\frac{1}{\Gamma(d/2)}\right|_{d=4-\varepsilon} = \frac{1}{16\pi^2} + \frac{\varepsilon}{32\pi^2}\left(1-\gamma+\ln 4\pi\right) + \mathcal{O}(\varepsilon^2) \tag{5.64}$$

where we have used the recurrence relation $\Gamma(z+1) = z\,\Gamma(z)$ and asymptotic formula

$$\Gamma(-\varepsilon/2) = -\frac{2}{\varepsilon} - \gamma + \mathcal{O}(\varepsilon) \tag{5.65}$$

for the Gamma function as $\varepsilon\to 0$, where γ is the Euler–Mascheroni constant $\gamma\approx 0.5772$. Since ε is small the first eigenvalue is negative, so the mass term ϕ^2 is a *relevant* perturbation of the Wilson–Fisher fixed point. On the other hand, the operator $-a(3+\varepsilon/2)\phi^2 + 2(3+\varepsilon)\phi^4$ corresponding to σ_4 corresponds to an *irrelevant* perturbation. The projection of RG flows to the (g_2, g_4) subspace is shown in figure 8.

Although we've seen the existence of the Wilson–Fisher fixed point only for $0<\varepsilon\ll 1$, more sophisticated techniques can be used to prove its existence in both $d=3$ and $d=2$ where it in fact corresponds to the Ising Model CFT. As shown in figure 8, both the Gaussian and Wilson–Fisher fixed–points lie on the critical surface, and a particular RG trajectory emanating from the Gaussian model corresponding to turning on the operator σ_4

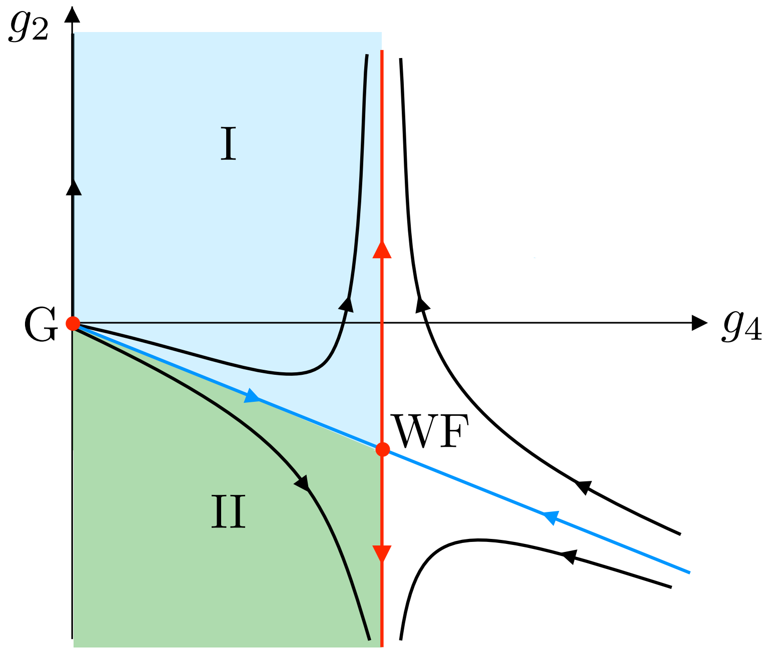

Figure 8: *The RG flow for a scalar theory in three dimensions, projected to the* (g_2, g_4) *subspace. The Wilson–Fisher and Gaussian fixed points are shown. The blue line is the projection of the critical surface. The arrows point in the direction of RG flow towards the IR.*

ends at the Wilson–Fisher fixed point in the IR. Theories on the line heading vertically out of the Gaussian fixed–point correspond to massive free theories, while theories in region I are massless and free in the deep UV, but become interacting and massive in the IR. These theories are parametrized by the scalar mass and by the strength of the interaction at any given energy scale. Theories in region II are likewise free and massless in the UV but interacting in the IR. However, these theories have $g_2 < 0$ so that the mass term is negative. This implies that the minimum of the potential $V(\phi)$ lies away from $\phi = 0$, so for theories in region II, ϕ will develop a vacuum expectation value, $\langle\phi\rangle \neq 0$. The RG trajectory obtained by deforming the Wilson–Fisher fixed point by a mass term is shown in red. All couplings in any theory to the right of this line diverge as we try to follow the RG back to the UV; these theories do not have well–defined continuum limits.

5.3.5 Zamolodchikov's C–theorem

Polchinski's equation showed that renormalization group flow could be understood as a form of heat flow. It's natural to ask whether, as for usual heat flow, this can be thought of as a *gradient* flow so that there is some real positive function $C(g_i, \Lambda)$ that decreases monotonically along the flow. Notice that this implies $C = \text{const.}$ at a fixed point g_i^*, and that $C(g_i^*, \Lambda) > C(g_i^{**}, \Lambda')$ whenever a fixed point g_i^{**} may be reached by perturbing the theory a fixed point g_i^* by a relevant operator and flowing to the IR. In 1986, Alexander Zamolodchikov found such a function C for any unitary, Lorentz invariant QFT in two dimensions.

Consider a two dimensional QFT whose (improved) energy momemtum tensor is given by $T_{\mu\nu}(x)$. This is a symmetric $2{\times}2$ matrix, so has three independent components. Introducing complex coodinates $z = x_1 + \mathrm{i}x_2$ and $\bar{z} = x_1 - \mathrm{i}x_2$, we can group these components as

$$\begin{aligned} T_{zz} &:= \frac{\partial x^\mu}{\partial z}\frac{\partial x^\nu}{\partial z}\,T_{\mu\nu} = \frac{1}{2}\,(T_{11} - T_{22} - \mathrm{i}T_{12}) \\ T_{\bar{z}\bar{z}} &:= \frac{\partial x^\mu}{\partial \bar{z}}\frac{\partial x^\nu}{\partial \bar{z}}\,T_{\mu\nu} = \frac{1}{2}\,(T_{11} - T_{22} + \mathrm{i}T_{12}) \\ T_{z\bar{z}} &:= \frac{\partial x^\mu}{\partial z}\frac{\partial x^\nu}{\partial \bar{z}}\,T_{\mu\nu} = \frac{1}{2}\,(T_{11} + T_{22}) \end{aligned} \tag{5.66}$$

where $T_{\bar{z}z} = T_{z\bar{z}}$. This stress tensor is conserved, with the conservation equation being

$$0 = \partial^\mu T_{\mu\nu} = \partial_{\bar{z}} T_{zz} + \partial_z T_{\bar{z}\bar{z}} \tag{5.67}$$

in terms of the complex coordinates. Note that the stress tensor is a smooth function of z and $\bar{z}$.

The two–point correlation functions of these stress tensor components are given by

$$\begin{aligned} \langle T_{zz}(z,\bar{z})\,T_{zz}(0,0)\rangle &= \frac{1}{z^4}F(|z|^2) \\ \langle T_{zz}(z,\bar{z})\,T_{z\bar{z}}(0,0)\rangle &= \frac{4}{z^3\bar{z}}G(|z|^2) \\ \langle T_{z\bar{z}}(z,\bar{z})\,T_{z\bar{z}}(0,0)\rangle &= \frac{16}{|z|^4}H(|z|^2) \end{aligned} \tag{5.68}$$

where the explicit factors of z and $\bar{z}$ on the *rhs* follow from Lorentz invariance, which also requires that the remaining functions F, G and H depend on position only through $|z|$. Like any correlation function, these functions will also depend on the couplings and scale Λ used to define the path integral.

The two–point function $\langle T_{z\bar{z}}(z,\bar{z})\,T_{z\bar{z}}(0)\rangle$ appearing here satisfies an important positivity condition. Using canonical quantization, we insert a complete set of QFT states to find

$$\begin{aligned} \langle T_{z\bar{z}}(z,\bar{z})\,T_{z\bar{z}}(0)\rangle &= \sum_n \langle 0|\hat{T}_{z\bar{z}}(z,\bar{z})\,\mathrm{e}^{-H\tau}|n\rangle\,\langle n|\hat{T}_{z\bar{z}}(0,0)|0\rangle \\ &= \sum_n \mathrm{e}^{-E_n\tau}|\langle n|\hat{T}_{z\bar{z}}(0,0)|0\rangle|^2 \end{aligned} \tag{5.69}$$

so that this two–point function is positive definite, and it follows that $H(|z|^2)$ is also positive definite.

Zamolodchikov now used a combination of this positivity condition and the current conservation equation to construct a certain quantity $C(g_i,\Lambda)$ that decreases monotonically along the RG flow. In terms of the two–point functions, current conservation (5.67) for the energy momentum tensor becomes

$$4F' + G' - 3G = 0 \qquad \text{and} \qquad 4G' - 4G + H' - 2H = 0 \tag{5.70}$$

where $F' = dF(|z|^2)/d|z|^2$ *etc.* We define the C**-function** to be

$$C(|z|^2) := 2F - G - \frac{3}{8}H \tag{5.71}$$

which obeys $dC/d|z|^2 = -3H/4$ by the current conservation equations. But by the positivity of the two–point function $\langle T_{z\bar{z}}(z,\bar{z})\, T_{z\bar{z}}(0)\rangle$ this means that

$$r^2 \frac{dC}{dr^2} < 0 \tag{5.72}$$

so C decreases monotonically under two dimensional scaling transformations, or equivalently under two dimensional RG flow. The value of C at an RG fixed point can be shown to be the **central charge** of the CFT.

Ever since it was first proposed, physicists have searched for a generalization of Zamolodchikov's theorem to RG flows in higher dimensions. The two–dimensional quantity $T_{z\bar{z}}$ is just the trace T^{μ}_{μ} of the energy momentum tensor and in 1988 John Cardy proposed that a certain term – known as "a" — in the expansion of the two–point correlator of T^{μ}_{μ} plays the role of Zamolodchikov's C in any even number of dimensions. Cardy's conjecture was verified to all orders in perturbation theory the following year by DAMTP's own Hugh Osborn, while a complete, non–perturbative proof was finally given in 2011 by Zohar Komargodski & Adam Schwimmer.