https://doi.org/10.1038/s41567-023-02300-3

Always relevant

It has been around fifty years since Kenneth Wilson's work on the renormalization group. *Nature Physics* celebrates this anniversary with a collection of Comments on its development and applications.

For the most part, studying physics is a continuous process of specialization. As part of postgraduate study, or the late part of undergraduate education, we choose specific topics that lead us into different fields. This often results in physicists losing touch with other areas and, in some cases, in the lack of mutual understanding between specialists.

The renormalization group is a technique in theoretical physics that bucks this trend. Typically a graduate level topic, it provides tools to understand how physics changes at different length scales. Whether working to identify which microscopic laws are responsible for the world around us, or what can emerge from the laws we already know, the renormalization group is a language that is shared between seemingly disparate fields, such as particle physics and condensed-matter physics.

The modern formulation of the renormalization group relies heavily on work by Kenneth Wilson in the 1970s, for which he was awarded the Nobel Prize[1]. In this Focus issue, we celebrate the 50th anniversary of his work on the topic with a group of Comments.

The collection begins with a piece from Philip Phillips reviewing the evolution of the renormalization group from its origins in the 1930s to potential progress beyond the Wilsonian framework. Much of the work — including that of Wilson — took place during the Cold War, a time of fundamental developments in many-body theory but also of significant struggles in international collaborations. Nevertheless, Soviet physicists had an important influence on Wilson's work, as described by Premala Chandra.

The next four pieces describe different applications of this theoretical technique. For example, a challenge in finding a quantum theory of gravity is that perturbation theory cannot be applied in the same way as the other fundamental forces. Astrid Eichhorn surveys the role of nonperturbative renormalization group techniques in potentially identifying a microscopic theory of quantum gravity.

More abstractly, Jaewon Song discusses the use of renormalization group tools to explore the space of quantum field theories. When these tools are combined with supersymmetry, it becomes possible to identify and study strongly coupled theories for which more common perturbative methods no longer work.

The renormalization group also makes close contact with current experiments. As Diogo Boito explains in the context of quantum chromodynamics, renormalization group methods enhance the precision of perturbative calculations used to test the standard model of particle physics.

Moving to larger length scales, Yuhai Tu discusses the application of the renormalization group to non-equilibrium models and especially biological systems. He recounts the discovery that a two-dimensional model of flocking features long-range order at finite temperatures — a behaviour that is impossible in equilibrium low-dimensional systems.

Many tools of modern theoretical physics were developed using a degree of physical intuition often disregarding mathematical rigour. This includes the techniques involved in many renormalization group studies. In our final Comment, Antti Kupiainen summarizes efforts by mathematical physicists to put renormalization group schemes on a formal footing.

It would be impractical to cover all the progress and applications of the renormalization group since the 1970s, and we did not attempt to do so. For example, the collection does not cover complex networks, where the formulation of renormalization methods is still an open problem[2]; or the density matrix renormalization group, which has provided many important advances in the understanding of quantum chemistry and many-body systems[3].

One of the striking implications of the renormalization group is that phase transitions in many different models and physical systems can be classified in a relatively small group of universality classes. Interactions can be classed as relevant, irrelevant and marginal, depending on their significance at large length scales.

This categorization can give the impression of a research programme that provides a definitive answer and comes to an end. The broad — and growing — influence of the renormalization group instead suggests that it will always be relevant.

Published online: 9 November 2023

References

1. Wilson, K. G. Nobel Lecture: "The renormalization group and critical phenomena." Nobel Lectures 1981–1990. (ed Ekspöng, G.) (World Scientific, Singapore, 1993).
2. Villegas, P. et al. *Nat. Phys.* **19**, 445–450 (2023).
3. Verstraete, F. et al. *Nat. Rev. Phys.* **5**, 273–276 (2023).

https://doi.org/10.1038/s41567-023-02208-y

Fifty years of Wilsonian renormalization and counting

Philip W. Phillips

Renormalization began as a tool to eliminate divergences in quantum electrodynamics, but it is now the basis of our understanding of physics at different energy scales. Here, I review its evolution with an eye towards physics beyond the Wilsonian paradigm.

Despite their microscopic differences, all simple fluids undergo a transition to the gas phase with identical universal characteristics. By systematizing the underpinnings of this universality, Kenneth Wilson formulated a far-reaching renormalization group (RG) principle[1], and in so doing established the tools for the modern understanding of phase transitions, critical phenomena and quantum field theory.

Pre-Wilson field theory

Before Wilson tackled the question of universality[1], quantum field theory had been developed through efforts to combine quantum physics with special relativity. However, this introduced the problem of vacuum polarization.

Since the 1930s[2] it had been known that the interaction of electromagnetic fields with the continuous distribution of 'negative energy' states (positrons) amends Coulomb's law with a logarithmic divergence to linear order in the fine-structure constant. At short distances, the divergence obtains $r \ll \hbar/m_e c = 3.86 \times 10^{-13} m_e$, where m_e is the electron mass and c the speed of light.

Fortunately, this divergence can be eliminated by defining a new effective charge, which will depend on the energy scale. It is from this dependence that the idea of a 'running' coupling constant emerges. Murray Gell-Mann and Francis Low showed[3] that to all orders in the fine-structure constant, the vacuum polarization at energy scales μ that are large relative to the mass of an electron, modifies the coupling constant $g(\mu)$ in accordance with the scale-invariant form:

$$\psi(g(\mu)) = \psi(g(\mu'))\left(\frac{\mu}{\mu'}\right)^a, \tag{1}$$

where a is a number and ψ is some function; neither are important for this discussion. This result shows that as the energy scale is varied, the new coupling constant is related to the original one by a scale-invariant or self-similar scale factor, $(\mu/\mu')^a$. Considering μ and μ' as infinitesimally separated leads to a differential equation that in its modern form is written:

$$\frac{dg}{d\ln\mu} = \beta(g). \tag{2}$$

Years before this equation was derived, Heisenberg noted that the fine-structure constant $\alpha \approx 1/137 \approx 2^{-4}3^{-3}\pi$ to an accuracy of 10^{-4}. The essence of Eq. (2) is that it is pointless to ruminate over any particular value for α. Instead, because of the charge renormalization, the fine-structure constant depends on the energy scale at which it is measured, typically represented by the momentum transferred by the interaction.

In pure quantum electrodynamics (QED) consisting of a single photon field and an electron, the solution to Eq. (2) predicts that the effective fine-structure constant,

$$\alpha_{eff} = \frac{\alpha}{1 - \frac{\alpha}{3\pi}\ln\frac{-q^2}{e^{5/3}m_e^2}}, \tag{3}$$

depends explicitly on the transferred momentum, q, where $-q^2 > 0$ is an increasing function of energy. This behaviour was directly observed in the Large Electron-Positron (LEP) collider in 1994. Although a full treatment with all the leptons and quarks is necessary to obtain the complete flow of α_{eff} from $\alpha_{eff}(M_W^2) = 1/128$ (M_W the mass of the W-boson) to its low-energy value of approximately 1/137, Eq. (3) is sufficient to capture the deviation from the naive expectation that the local quasi-instantaneous physics and hence only the bare parameters in the Lagrangian should matter in the high-energy limit. This is not borne out in QED. Fig. 1 depicts that quantum chromodynamics (QCD) – the theory of strong interactions between gluons and quarks – stands in contradistinction to QED. This is one of the great triumphs of Wilson's renormalization approach[4].

Although a theory with photon fields is naturally scale-invariant, QED tells us that once matter is included, such scale invariance is lost by virtue of the running of the charge manifested in Eq. (3). Nonetheless, the presence of a logarithm in the β function reflects, to quote Wilson, "a problem lacking a characteristic scale"[1]. In fact, a similar logarithm arises in the theory for the ground-state energy of an electron gas, which features a Fermi sea of positive-energy electron states rather than the negative-energy positrons of the vacuum.

How are these two features of QED compatible? In QED and elementary particle theory in general, the only discernible energy scale is set by the rest mass of the constituents. Integrals of the form

$$\int_{m_e c^2}^{\infty} \frac{dE}{E} \tag{4}$$

are logarithmically divergent precisely because all energy scales above $m_e c^2$ contribute equally. Consider a scale $E' > mc^2$. The contribution to the integral from E' to $2E'$ is simply ln2, independent of the scale E'.

In practice, all field theories are defined up to a high-energy cutoff or equivalently a short-distance scale. Precisely the role played by the high-energy (short-distance) cutoff in an analysis of field theories lies at

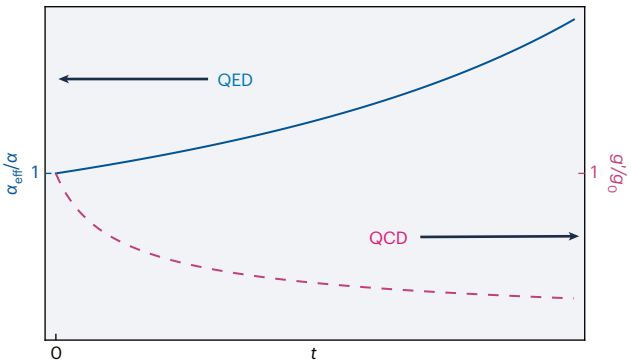

Fig. 1 | Running of the coupling constants. Illustrative plot of the β functions for the coupling constants in QED, Eq. (3), and QCD, Eq. (13), as a function of the energy scale, t. While both flow under renormalization, they do so in opposite directions. QED becomes more strongly coupled at high energy while QCD does just the opposite. At high energy, QCD is asymptotically free as the coupling constant vanishes. Confinement of the basic constituents, quarks and gluons, obtains at low energy in QCD.

the heart of Wilson's approach to renormalization. As we will show, what Wilson clarified is that low-energy theories depend on short-distance physics through operators classified as relevant, marginal and, in some cases, irrelevant depending on the energy scale being probed. It is from this dependence that universality arises.

To set this up, we note that in field theory it is not the value of a field at any point that matters, but rather correlation functions of the underlying fields. A key precursor to Wilson's work was the Callan-Symanzik[5,6] equation,

$$\left[\mu\frac{\partial}{\partial\mu}+\beta(g)\frac{\partial}{\partial g}+n\gamma(g)\right]G^n(\mu,g,\gamma)=0, \tag{5}$$

which established that any n-point correlation function G is independent of the cutoff through two universal functions that communicate the shift in the coupling constant, $\beta(g)$, and the field strength, $\gamma(g)$, in such a way to counteract the shift made in the energy scale, μ.

Block renormalization

The story of renormalization thus far, prior to 1971, is more tied to removing infinities that arise in computing Feynman graphs than it is to some universal physical principle involving collective degrees of freedom. Wilson provided[1] this missing link by focusing on how systems with fluctuations on all length scales, such as a boiling pot of water, can be studied without forgoing locality.

One of the simplest models featuring a phase transition is the Ising model for the onset of ferromagnetism. In this model, spins with either an up or down degree of freedom occupy sites with a separation of a on a d-dimensional lattice and interact with nearest-neighbour interactions. In this context, Leo Kadanoff introduced a block coarse-graining renormalization scheme[7] for the Ising model in which the entire system is divided into cells of edge length ba ($b>1$). This approach provides an operational way to build in fluctuations smaller than the correlation length, ξ.

A new coarse-grained spin variable is introduced to represent the average of the b^d spins in each block. The Hamiltonian can then be rewritten to take the same form at each iteration as long as the block spins are normalized by a rescaling factor to maintain the up-down or Z_2 Ising symmetry.

The major conceptual leap in this approach is the assumption that the blocks, like the underlying spins, only have nearest-neighbour interactions. An initial system of N spins has Nb^{-d} effective spins after blocking, each separated by ba. The correlation length ξ can be represented either in units of the initial lengthscale $\xi=\xi_1\times a$ or the blocked lengthscale $\xi=\xi_b\times ba$. The rescaled correlation length ξ_b is smaller than the correlation length at the initial scale ξ_1:

$$\xi=\xi_b\times(ba)=\xi_1\times a\Rightarrow\xi_b=\frac{\xi_1}{b}. \tag{6}$$

Consequently, the corresponding rescaled Hamiltonian for the bth iteration H_b lies further away from a critical point – where the correlation length diverges – than the initial Hamiltonian H_1.

This is reflected by the rescaled temperature and magnetic field parameters in the model, t_b and h_b, respectively. Let t and h, be the bare values of the temperature and magnetic field, respectively. A key assumption in the renormalization group procedure is that after rescaling, these quantities satisfy power-law scaling laws, $t_b=tb^{y_t}$ and $h_b=hb^{y_h}$ where y_t and y_h are both positive and can only be determined from the full renormalization transformations. This leads to a series of recursion equations that ultimately make it possible to sample the infinite hierarchy of fluctuations with only a finite number of degrees of freedom at each step.

Power counting

A revolution came with Wilson's momentum-space translation[1] of the Kadanoff real-space coarse graining[7]. It represented the degrees of freedom in the Ising model as fields in continuous space. This approach brought the physics of critical phenomena into quantum field theory, and through renormalization established what field theory looks like in the statistical continuum limit.

The notion of renormalizabilty is in general ill-posed as normally stated, as one must mention the space of operators within which a theory is renormalizable. More explicitly, consider a certain theory described by a classically local action $S(\phi_i)$ of some fields $\phi_1,\cdots,\phi_n$. Suppose the field theory is valid up to some energy scale E_0 and we seek a theory valid for energies below this scale, $E<E_0$. To do this, one introduces a cutoff scale $\Lambda<E_0$ and 'integrates out' fields whose energy is higher than Λ to obtain an effective action S_Λ that depends only on low-energy degrees of freedom. This is the energy- and momentum-space equivalent process to the blocking step of Kadanoff's procedure.

Operationally, this is done by splitting the field into high and low-energy components

$$\phi(\omega)=\begin{cases}\phi_{\mathrm{L}}(\omega), & \omega<\Lambda\\ \phi_{\mathrm{H}}(\omega), & \omega>\Lambda\end{cases},$$

where ω is the energy, and performing an integration over the high-energy (H) modes in the partition function to obtain the effective low-energy (L) theory:

$$\int D\phi e^{iS(\phi)}=\int D\phi_{\mathrm{L}}e^{iS_\Lambda(\phi_L)} \tag{7}$$

where

$$S_\Lambda(\phi_L) = -i\ \log(\int D\phi_H e^{iS(\phi_L,\phi_H)}) \tag{8}$$

is the outcome of the integration.

In the analysis of running coupling constants there are special values known as fixed points for which $\beta = 0$ in Eq. (2). Defining S_* as the action at a particular fixed point, one can write the action for a different set of parameters as

$$S_\Lambda = S_* + \int d^d x \sum_i g_i \mathcal{O}_i \tag{9}$$

for some set of field operators $\mathcal{O}_i$ that are local despite the integration of high-frequency fields, because we focus on fields with $\omega < \Lambda$.

As with the block renormalization approach, we can consider the behaviour of the model under length rescaling,

$$x^\mu \to \lambda^{-1} x^\mu. \tag{10}$$

If, under such a transformation, an operator $\mathcal{O}(x)$ can be written as

$$\mathcal{O}(x) = \lambda^{d_\mathcal{O}}\, \mathcal{O}(\lambda^{-1}x), \tag{11}$$

we interpret $d_\mathcal{O}$ as the dimension of $\mathcal{O}$.

Under a rescaling, the action can be organized based on the exponent of λ in each term, a procedure known as power counting. In the $\lambda \to \infty$ limit, each operator will either remain invariant, vanish or diverge. The rule is as follows. Because of the d-dimensional spacetime measure in the action, operators with $d_\mathcal{O} - d > 0$ are irrelevant and do not influence the low-energy physics. Relevant operators correspond to $d_\mathcal{O} - d < 0$.

Operators with $d_\mathcal{O} - d = 0$ are marginal. In these cases, all scales are important and such operators are the origin of logarithms in the β function.

The core of renormalization is in the observation that there is a dimension D above which the operators are irrelevant. Furthermore, the number of local operators O_i whose dimension is less than (or equal) to D is finite. This obtains because classically local operators are polynomials in the fields ϕ and their derivatives. Since there are finitely many of these, one can make sense of such theories. Wilsonian renormalization rests on the simple principle that the low-energy physics is determined only by the relevant or marginal interactions, or in rare cases, irrelevant couplings but only at low enough scales. That the details of renormalization are determined by the dimension of operators rather than the nature of the microscopic features of the interactions or the cutoff is the origin of universality in the Wilsonian approach.

There are subtleties[8] in evaluating S_Λ, which typically has to be performed perturbatively. However, these can be overcome by a slight recasting[8] of the problem set forth by Wilson. We can imagine integrating out high-energy modes one small energy slice at a time. First we remove the modes with energies in the range $\Lambda > \omega > \Lambda - d\Lambda$, then $\Lambda - d\Lambda > \omega > \Lambda - 2d\Lambda$ and so on. At each stage the effective action S_Λ changes, which is described by the Wilson equation,

$$\frac{\partial S_\Lambda}{\partial \Lambda} = F(S_\Lambda), \tag{12}$$

where F is a well-defined functional that can be calculated.

As the Wilson equation represents a flow in an infinite dimensional space, examining its properties for a range of operators can be accomplished entirely from the eigenvalue spectrum. Irrelevant operators correspond to negative eigenvalues, which represent benign converging flows. If the functional is linearized around zero-coupling, the eigenvalues are precisely the numbers $d_\mathcal{O} - d$ obtained from power counting. As $F(S_\Lambda)$ is a smooth function of the couplings, there is no place[8] for singularities to obtain especially since we are performing a path integral over a narrow range of energy with both a low- and high-energy cutoff. Hence, if an eigenvalue is negative in the free theory, the same holds for the interacting theory. Power counting then rules even if the dimension can change at strong coupling, for example in the Thirring model; hence the claim of marginality or relevance is the crux of the matter.

The β function

The evolution of the action as high momentum states are integrated out is precisely what is described by the running of the coupling constants in the β function. What Wilson added beyond the Gell-Mann/Low flow equation, Eq. (2), is that the β function is governed by power-counting, coupled with integration of the high-energy modes and rescaling.

In the theory of QCD, perturbative treatment of non-Abelian Yang-Mills gauge theories[4,9] yields a β function of the form

$$\beta(g) = -bg^3 \to g'^2 = \frac{g_0^2}{1 + 2bg_0^2 t}, \tag{13}$$

where t is proportional to the energy transferred and b is a numerical constant. At high energies, $t \to \infty$, the coupling constant, g, flows to zero, producing the phenomenon known as asymptotic freedom whereby quarks and gluons become weakly interacting and treatable using perturbation theory. The opposite is true at low energies, where instead confinement of quarks and gluons takes place, producing a divergence of the coupling constant and the general breakdown of the whole perturbative scheme[4,9].

Similar phenomena occur for the seemingly unrelated problem of a localized magnetic spin engaging in spin-flip scattering with a non-interacting band of conduction electrons, which is known as the Kondo problem. The spin-flip scattering operator is marginal, and the coupling strength flows from an initial value of g_0 according to the β function

$$\beta(g) = g^2 \to g' = \frac{g_0}{1 - g_0 \ln E_0/E}, \tag{14}$$

where E_0 and E are the initial and final energy scales, respectively. If the initial interaction is antiferromagnetic, $g_0 < 0$, the flow is towards strong coupling, yielding a divergent coupling constant signalling the formation of a bound state between the impurity and the conduction electrons.

In both QCD and the Kondo problem, the formation of new degrees of freedom at low energies is obtained through a cross-over rather than a phase transition. A key triumph of Wilson's treatment of the Kondo problem is that it captured the universal scaling that transpires below a characteristic temperature T_k at the $g \to -\infty$ fixed point, where the local moment is completely screened. As fixed points are characterized by scale invariance, the Kondo temperature is obtained by imposing

the scale-invariant condition $\tilde{D}\partial T_k(g)/\partial\tilde{D} = 0$, where $\tilde{D}$ is the bandwidth of the conduction electrons. The universal physics of all properties such as resistivity, magnetism and thermodynamics below the cross-over scale T_k is a consequence of this scale-invariant condition at the strongly coupled fixed point of the theory.

Beyond Wilson

Fifty years on, it is natural to wonder if there is any physics beyond the Wilsonian paradigm. This could arise from a theory in which as the high energy is probed features emerge in the scattering matrix that are distinct from the poles that correspond to particles. For example, in the case of a doped Mott insulator[10] scattering matrix zeros describing non-propagating or incoherent degrees of freedom have been identified.

Before we get to the Mott problem, consider quantum gravity. As pointed out previously[11], probing high energy in a gravitational theory should produce black hole information, which cannot be represented as simple poles in the scattering matrix. This would lead to a failure of the Wilsonian separation of energy scales and provide an example of ultraviolet/infrared (UV/IR) mixing, which is well studied in the context of non-commutative field theories[12]. Precisely how such UV/IR mixing plays out in a theory of quantum gravity remains an open question.

So far, our best understanding of quantum gravity stems from gauge–gravity duality, also known as the AdS/CFT conjecture[13]. But even this conjecture has an effective Wilsonian interpretation at its core arising from the locality of energy in the β function. The central claim of the gauge–gravity duality is that some strongly coupled conformally invariant field theories in d dimensions are dual to a theory of gravity in a $d+1$ spacetime that is asymptotically described by an anti de Sitter (AdS) metric parameterized by

$$\mathrm{d}s^2 = \frac{R^2}{z^2}\left(\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{d}z^2\right), \tag{15}$$

where R and z are the radius and radial coordinates of the AdS spacetime, respectively. This spacetime is invariant under the transformation $x_\mu \to \Lambda x_\mu$ and $z \to z\Lambda$ and hence satisfies the requisite symmetry for the implementation of the gauge–gravity duality, although not the full symmetry of the conformal group. The conformal field theory is viewed as lying on the $z = 0$ boundary of the AdS spacetime.

Our current understanding of the radial coordinate z is that it represents the flow in the energy scale during renormalization. The scale change, $x_\mu \to \Lambda x_\mu$ increases the radial coordinate, $z \to z\Lambda$. Consequently, moving towards greater z in the bulk of the geometry increases the corresponding projection onto the boundary, as depicted in Fig. 2. The limit of $z = \infty$ therefore represents the full low-energy or IR limit of the strongly coupled theory. The AdS/CFT conjecture thereby provides a complete geometrization of the renormalization group procedure.

The second area where a possible breakdown of the Wilsonian paradigm might arise is the strange metal[14] phase in the cuprate superconductors, which all start out as Mott insulators. In the strange metal, the resistivity increases way beyond the limit set by a scattering length determined by the physics of the underlying lattice constant. As such a length scale determines the cutoff for particle scattering, the strange metal with its non-saturating resistivity requires physics beyond the Fermi liquid quasiparticle picture. In terms of the standard field theory for a Fermi surface[8,15], no operator in any effective Lagrangian exists that can account for a non-saturating resistivity from the lowest temperatures to temperature scales where the particle-picture breaks down. This physics ultimately arises from the incoherent part of the spectrum, that is, zeros of the scattering matrix, which has no Wilsonian formulation at present. So indeed, these two examples indicate that much physics possibly lies beyond the Wilsonian paradigm.

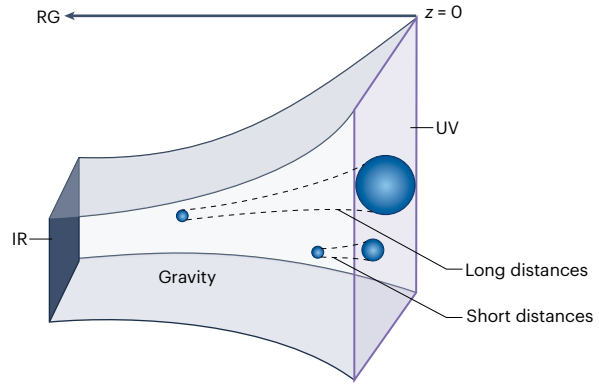

Fig. 2 | Geometrical representation of the key claim of the gauge–gravity duality. A strongly coupled field theory lives at the boundary, at the high-energy UV scale. The horizontal z-direction (the extra dimension in the gauge–gravity duality) represents the running of the renormalization scale. This is illustrated by the two projections at different values of z in the spacetime. Because the spacetime is asymptotically hyperbolic, larger values of z lead to a larger projection of the boundary theory and a full running of the renormalization scale amounts to the construction of the low-energy, IR limit of the original strongly coupled UV theory.

Philip W. Phillips
Department of Physics and Institute of Condensed Matter Theory, University of Illinois at Urbana-Champaign, Urbana, IL, USA.
e-mail: dimer@illnois.edu

Published online: 9 November 2023

References

1. Wilson, K. G. *Rev. Mod. Phys.* **47**, 773 (1975).
2. Serber, R. *Phys. Rev.* **48**, 49 (1935).
3. Gell-Mann, M. & Low, F. E. *Phys. Rev.* **95**, 1300 (1954).
4. Gross, D. J. & Wilczek, F. *Phys. Rev. Lett.* **30**, 1343 (1973).
5. Callan, C. G. Jr *Phys. Rev. D* **2**, 1541 (1970).
6. Symanzik, K. *Commun. Math. Phys.* **18**, 227 (1970).
7. Kadanoff, L. P. *Physics Physique Fizika* **2**, 263 (1966).
8. Polchinski, J. Preprint at https://arxiv.org/abs/hep-th/9210046 (1992).
9. Politzer, H. D. *Phys. Rev. Lett.* **30**, 1346 (1973).
10. Dzyaloshinskii, I. *Phys. Rev. B* **68**, 085113 (2003).
11. Berglund, P. et al. Preprint at https://arxiv.org/abs/2202.06890 (2022).
12. Seiberg, N. & Witten, E. *J. High Energy Phys.* **1999**, 032 (1999).
13. Maldacena, J. *Int. J. Theor. Phys.* **38**, 1113 (1999).
14. Phillips, W., Hussey, N. E. & Abbamonte, P. *Science* **377**, eabh4273 (2022).
15. Shankar, R. *Rev. Mod. Phys.* **66**, 129 (1994).

Acknowledgements

This work was funded through NSF DMR-2111379. I thank R. Leigh and G. La Nave for always insightful conversations and J. Zhao for help with Fig. 1.

Competing interests

The author declares no competing interests.

https://doi.org/10.1038/s41567-023-02191-4

Soviet influences on Kenneth Wilson's renormalization group work

P. Chandra

Kenneth Wilson worked on the renormalization group during the Cold War, when communication between scientists in the Soviet Union and in the West was restricted. Nevertheless, Soviet physicists had a strong influence on Wilson's work.

Despite the barriers imposed by the Cold War, Kenneth Wilson's work on renormalization was significantly influenced by physics from the Soviet Union, as he acknowledged on several occasions. In his Nobel Lecture[1], Wilson cited several independent Soviet efforts[2–10] in the pre-1971 period that were closely related to his own renormalization work. Indeed Wilson made several trips to Moscow around 1970[11–14], and the fact that he did so frequently (Fig. 1) suggests that he found these visits very beneficial.

Already as a young researcher, Wilson was influenced by Soviet physicists; he later admitted[11] to learning about renormalization in quantum electrodynamics[15] from the book of Nikolay Bogoliubov and Dmitry Shirkov[16], which is said to have made his thesis advisor, Murray Gell-Mann, quite annoyed. The aim of this Comment is to convey a flavor for the Soviet influences on Wilson's renormalization group work with references for the interested reader who would like to pursue more detail.

Fig. 1 | Bertrand Halperin, Kenneth Wilson and Mark Azbel in Moscow in 1977. Wilson continued to visit Moscow after the 69–70 trips discussed in this Commentary and he spoke at a 'Sunday seminar' for 'refuseniks'[32] in 1977. Image courtesy of James S. Langer.

Soviet developments

It is reported that Lev Landau himself considered his mean field theory of phase transitions to be incomplete, as it could not describe ordered systems with significant fluctuations[13,17,18]. In addition, the exact solution of the two-dimensional Ising model gives singular thermodynamic behaviour[19]. This cannot be reproduced using Landau theory.

In the early 1960s Alexander Voronel and his colleagues at the Institute of Physical and Technological Measurements in Moscow investigated the specific heat of argon near its critical point. They discovered an anomaly that bore striking similarity to that observed earlier at the superfluid transition of liquid helium[20]. This removed the possibility that the critical point behaviour in helium was a quantum effect.

Voronel shared his results with many researchers including several abroad[21–23]. This was a bold step at a time when scientific exchange between the Soviet Union and the West was restricted. Voronel had been arrested at age fourteen for his political activities[22] and was known to support dissidents[24]; he was thus most probably on a Soviet watch list. Still Voronel was eager to tell the international community about his findings. Michael Fisher recognized their importance immediately and requested Voronel's numerical data[21–23]. Here the influence of the Cold War is evident: it seems that the Fisher–Voronel correspondence was compromised as these researchers did not receive all of each other's letters[22]. To prevent further loss of information, in a break with the usual protocol of the time, Voronel included the requested numerical data at the end of a journal article so that it would be publicly accessible[22]. In later years Fisher often showed Voronel's specific heat singularity in his talks as a key motivation for the study of classical critical phenomena[22].

At roughly the same time, Jan Sengers in the Netherlands was also challenging the conventional van der Waals–Landau approach to criticality with his transport measurements[21]. These two sets of experiments contributed to the growing collective feeling everywhere that Gaussian fluctuations around mean-field theories were not enough to describe many classical critical phenomena[21].

Subsequent developments in the West during the 1960's have been well documented, particularly by Cyril Domb[25]. In this Comment the focus will therefore be on the lesser-known theoretical progress in classical criticality during that time period in the Soviet physics community.

According to Alexander Polyakov[12–14], the modern development of this subject started with the work of Alexander Patashinski and Valery Pokrovsky[2,3]. Their approach was inspired by a proposal by Landau in the late 1950s to express the partition function as a path integral where the Landau free energy emerges as a saddle-point solution[18].

Patashinski and Pokrovsky put forward the idea that the physics of the critical point is scale-invariant[2,3]. This produces scaling relations of the exponents describing the singularities of different thermodynamic functions at the critical point. It also explains universality, the emergence of identical singular behaviour in different systems, such as the cases of superfluid helium and argon. Similar ideas about critical phenomena were developed independently and roughly concurrently in the West by Fisher, Domb, Leo Kadanoff and their collaborators[26–28].

The results of Patashinski and Pokrovsky[2,3] convinced the Soviet physics community of universality, since they had shown that the problem of classical criticality does not depend on details of a system's short-range physics[12,14]. At the same time, work by Polyakov[4] and by Alexander Migdal[5] demonstrated the close connection between critical phenomena and relativistic quantum field theory.

The confluence of universality, quantum field theory and critical phenomena is apparent in a pioneering study of the four-dimensional Ising model by Anatoly Larkin and David Khmelnitskii[6]. It was already known that Landau theory breaks down for the Ising model in four or fewer dimensions, and that the model's universal behaviour is equivalent to ϕ^4 field theory. Exploiting this link, Larkin and Khmelnitskii applied the renormalization methods of quantum electrodynamics to the ϕ^4 model, finding clear singularities in the exponents of the specific heat and other quantities. Finally, they noted that the four-dimensional Ising model is realized in a three-dimensional uniaxial ferroelectric; here anistropic dipolar interactions effectively add an extra dimension[6]. This was the first exact calculation of a non-mean field exponent in an experimentally realizable system, with later measurements confirming the predictions[29,30].

Wilson's visits to the Soviet Union

Given their shared interests in critical phenomena and relativistic field theory, it is not surprising that Wilson visited his colleagues in the Soviet Union, particularly Migdal and Polykov[11], in 69–70 even though such trips were still quite unusual for US citizens at that time. Polyakov reflects that he and Migdal were very keen to learn more about Wilson's renormalization work, even though it was based on an approximate recursion scheme[13,14]. Polyakov writes[13]: "Trying to understand it, I derived it by some crude truncation of Feynman's diagrams. Ken liked the derivation (and generously included it in his later review[31], but I thought it just showed that the recursion formula was too primitive. However, later it helped Ken to develop a general approach to the renormalization group and epsilon expansion."

Philosophically, Wilson was not deterred by approximate expressions, particularly as he could solve them computationally; here he may well have been influenced by his father who was a theoretical chemist[11]. Polyakov notes that quantum field theory was considered by many in the Soviet high-energy community at that time to have pathological technical issues. It was thus refreshing for both Polyakov and Migdal to see that Wilson shared their belief in the natural connection between particle physics and critical phenomena[13,14]. Polyakov also comments[13]: "In spite of our different 'ideologies', I was very impressed by the power and depth of Ken's arguments, and learned lots of subtle things from our discussions."

During Wilson's later visits to Moscow, he also spoke at a 'Sunday seminar' for 'refuseniks'[32] organized by Voronel and his colleague Mark Azbel. Refusenik was the unofficial term for a person, typically a Soviet Jew, who was denied permission to emigrate, usually to Israel. Since that time the word refusenik has entered the colloquial English lexicon to mean a person who refuses to follow the law particularly as a form of protest. In the former Soviet Union, refuseniks usually lost their jobs, which for scientists meant exclusion from their community in all forms. In Moscow, Voronel and Azbel organized regular seminars to provide mutual support and intellectual sustenance for the refusenik scientists[24,32,33]. In 1977 Voronel and Azbel organized a meeting on collective phenomena in physics to mark the fifth anniversary of the Moscow seminar series, and Wilson was among the invited speakers from abroad. These international visitors were given a chilly welcome by the Soviet authorities, and indeed Wilson was detained at the Moscow airport for several hours after he stated the purpose of his visit[32].

Wilson clearly believed that scientific discussions with his Soviet colleagues contributed significantly to the development of his work on the renormalization group; furthermore he refused to be deterred by the tense Cold War relations between the Soviet Union and the West at the time. It seems fitting to end this Commentary with the words of Wilson himself at his Nobel Banquet[34]: "The hardest problems of pure and applied science can only be solved by the open collaboration of the world-wide scientific community. Scientists under all forms of government must be able to participate fully in international efforts."

These sentiments continue to be important and relevant today.

P. Chandra
Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, USA.
e-mail: pchandra@physics.rutgers.edu

Published online: 9 November 2023

References

1. Wilson, K. G. *The Renormalization Group and Critical Phenomena* (The Nobel Foundation, 1982).
2. Patashinskii, A. Z. & Pokrovskii, V. L. *Sov. Phys. JETP* **19**, 677 (1964). [*ZhETP* **46**, 994 (1964)].
3. Patashinskii, A. Z. & Pokrovskii, V. L. *Sov. Phys. JETP* **23**, 292 (1966). [*ZhETF* **50**, 439 (1966)].
4. Polyakov, A. M. *Sov. Phys. JETP* **28**, 533 (1969). [*ZhETF* **55**, 1026 (1968)].
5. Migdal, A. A. *Sov. Phys. JETP* **28**, 533 (1969). [*ZhETF* **55**, 1964 (1968)].
6. Larkin, A. I. & Khmelnitskii, D. E. *Sov. Phys. JETP* **29**, 1123 (1969). [*ZhETF* **56**, 2087 (1969)].
7. Polyakov, A. M. *Sov. Phys. JETP* **30**, 151 (1970). [*ZhETF* **57**, 271 (1969)].
8. Polyakov, A. M. *Sov. Phys. JETP* **32**, 296 (1971). [*ZhETF* **59**, 542 (1970)].
9. Midgal, A. A. *Sov. Phys. JETP* **32**, 552 (1971). [ZhETF **59**, 1015 (1970)].
10. Patashinskii, A. Z. & Pokrovskii, V. L. *Fluctuation Theory of Phase Transitions* (Pergamon Press, 1979).
11. Interview with Kenneth G. Wilson. *Physics of Scale* (6 July 2002).
12. Interview with Alexander M. Polyakov. *Physics of Scale* (6 February 2003).
13. Polyakov, A. M. Preprint at https://arxiv.org/abs/1502.03502 (2015).
14. Private Communication with A. M. Polyakov (2023).
15. Gell-Mann, M. & Low, F. E. *Phys. Rev.* **95**, 1300–1312 (1954).
16. Bogoliubov, N. N. & Shirkov, D. V. *Introduction to the Theory of Quantized Fields* (Interscience, 1959).
17. Interview with Alexander Z. Patashinski. *Physics of Scale* (10 July 2002).
18. Pokrovsky, V. L. *Ann. Rev. Condens. Phys.* **13**, 15 (2022).
19. Onsager, L. *Phys. Rev.* **65**, 117–149 (1944).
20. Bagatskii, A. V., Voronel, M. I. & Gusak, V. G. *Sov. Phys. JETP* **16**, 517 (1963). [*ZhETP* **43**, 728 (1962)].
21. Anisimov, M. A. *Int. J. Thermophys.* **32**, 2001 (2011).
22. Interview with Alexander V. Voronel. *Physics of Scale* (5 December 2002).
23. Fisher, M. E. *The Nature of Critical Points* (University of Colorado Press, 1965).
24. Charap, J. M. *Nature* **297**, 165 (1982).
25. Domb, C. *The Critical Point: A Historical Introduction to the Modern Theory of Critical Phenomena* (Taylor & Francis, 1996).
26. Essam, J. W. & Fisher, M. E. *J. Chem. Phys.* **38**, 802 (1963).
27. Domb, C. & Hunter, D. L. *Proc. Phys. Soc.* **86**, 1147 (65).
28. Kadanoff, L. P. *Physics* **2**, 263 (1966).
29. Ahlers, G., Kornblit, A. & Guggenheim, H. J. *Phys. Rev. Lett.* **34**, 1227–1230 (1975).
30. Sandvold, E. & Courtens, E. *Phys. Rev. B* **27**, 5660–5668 (1983).
31. Wilson, K. G. & Kogut, J. *Phys. Rep.* **12C**, 75 (1974).
32. Azbel, M. Ya. *Refusenik: Trapped in the Soviet Union* (Paragon House Publishers, 1981).
33. Rich, V. *Nature* **266**, 668 (1977).
34. *Kenneth G. Wilson's Speech at the Nobel Banquet, December 10, 1982* (The Nobel Foundation, 1982).

Acknowledgements

I am very grateful to B. L. Altshuler and to A. M. Polyakov for sharing their insights on theoretical developments of critical phenomena in the former Soviet Union with me. I also thank J. S. Langer for telling me about the Azbel seminars, and for contributing the photograph of Bertrand I. Halperin, Kenneth G. Wilson and Mark Ya. Azbel in Moscow in 1977. This work was supported by DOE Basic Energy Sciences grant DE-SC0020353; in its final stages, I was visiting the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452.

Competing interests

The author declares no competing interests.

https://doi.org/10.1038/s41567-023-02261-7

The microscopic structure of quantum space-time and matter from a renormalization group perspective

Astrid Eichhorn

The correct microscopic theory of quantum gravity may be an interacting, scale-invariant, 'asymptotically safe' model. This Comment discusses the renormalization group's role in defining asymptotic safety and understanding its consequences.

The tools of the renormalization group enable us to study theories on different length scales, including those far below experimental reach. They provide us with a mathematical analogue of a microscope to probe the structure of space-time and matter.

By applying renormalization group methods to a theory, we can understand how its interaction coupling constants 'flow' as the length scale is changed. Crucially, these renormalization group flows can feature fixed points at which the zooming of the renormalization group 'microscope' does not lead to changes.

Renormalization group fixed points

At a fixed point, the properties and response of a system are the same at all length scales. Scale symmetry of this kind provides a compelling paradigm for a fundamental symmetry. If this applies, then no new particles, dimensions or structures can exist at microscopic scales, because a scale-symmetric theory is self-similar and must look the same at all scales that are described by the fixed-point regime. However, the presence of quantum fluctuations can break scale symmetry and it is not guaranteed that a microscopic fixed point can be found, making scale symmetry challenging to achieve.

Scale symmetry at microscopic length scales can be achieved in two different ways. The first happens when the interactions in a theory tend to zero, which removes the effect of quantum fluctuations; this produces a fixed point that corresponds to a non-interacting model. This phenomenon is known as asymptotic freedom, which is an important feature of quantum chromodynamics and other theories in high-energy physics.

A second way happens when quantum fluctuations balance the finite coupling strengths. In this scenario, a genuine quantum version of scale symmetry, known as asymptotic safety, emerges. In asymptotically safe theories the fixed point of the renormalization group occurs at nonzero values of the couplings — an interacting fixed point.

Although the existence of an asymptotically safe fixed point has long been regarded as a somewhat exotic possibility in particle physics, condensed matter theory contains numerous examples. Interacting fixed points appear at second-order phase transitions, because these are characterized by a diverging correlation length. When this happens, there is no longer any distinct scale in the system so it must become self-similar and therefore correspond to a renormalization group fixed point. Because the couplings of such a system are generically nonzero, the fixed point must be an interacting one.

Many interacting fixed points have been identified in two- or three-dimensional quantum field theories, depending on the condensed-matter system at hand. Only recently have asymptotically safe fixed points been found in four dimensions[1].

The standard model of particle physics and renormalization group fixed points

Perhaps the most famous four-dimensional theory is the standard model of particle physics. There are a priori three possibilities for its microscopic behaviour: it may be asymptotically free, asymptotically safe or not 'ultraviolet complete' (which means it becomes invalid at small length scales).

Although some components of the standard model, such as quantum chromodynamics, are asymptotically free, this is not true of the whole theory. Because quantum fluctuations of charged-matter fields screen the coupling constant of the Abelian gauge interaction, the coupling constant decreases towards large length scales and cannot become asymptotically free. Asymptotic safety is harder to rule out because it can occur in strongly interacting regimes that are difficult to study theoretically. However, so far no evidence for asymptotic safety has been found and it appears that the standard model may not be an ultraviolet-complete theory without introducing new physics.

This theoretical argument for new physics is supported by observational arguments. For example, right-handed neutrinos were not part of the original standard model but must exist to explain neutrino oscillations. Other beyond-the-standard-model particles may exist, although the evidence is less definitive. For example, dark matter is often attributed to a new kind of particle, or there may be additional particles that play a part in the emergence of matter–antimatter asymmetry in the earlier Universe, a process known as baryogenesis.

However, another fundamental force is already known to exist beyond the three forces that are part of the standard model: gravity. We may therefore conjecture that a complete theory of the standard model coupled to quantum gravity flows from an interacting fixed point at short length scales to the known physics at large length scales. Testing this conjecture is fundamental to investigating whether gravity and matter can together be described by an asymptotically safe standard model.

Asymptotically safe gravity

Gravity is difficult to quantize. Perturbative methods produce increasing numbers of divergences, which require higher-order interaction terms to be introduced to cancel them, each of which comes with new

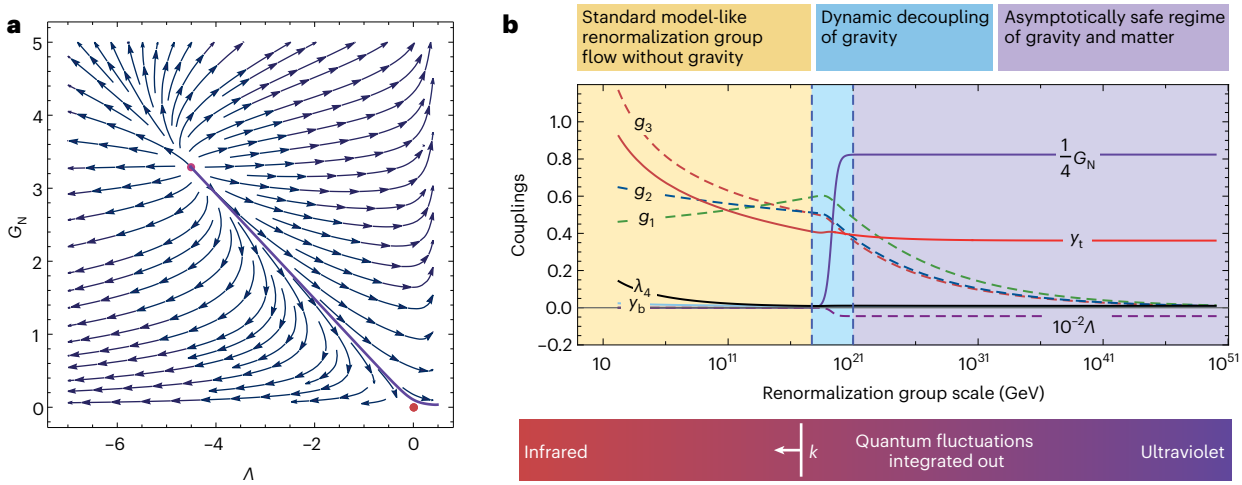

Fig. 1 | Renormalization group flow of gravity and standard model couplings. **a**, The renormalization group flow in the plane spanned by the dimensionless counterparts of the Newton coupling G_N and the cosmological constant Λ, which are the leading-order gravitational couplings. The asymptotically safe fixed point (magenta dot) has many renormalization group trajectories emanating from it. A special trajectory, which passes very close to the non-interacting fixed point (red dot), reproduces the measured values of the Newton constant and cosmological constant. **b**, The renormalization group flow of gravity, encoded in G_N and Λ, together with standard model couplings g_1 (Abelian gauge coupling), g_2 (SU(2) gauge coupling), g_3 (SU(3) gauge coupling), y_t (top quark Yukawa coupling), y_b (bottom quark Yukawa coupling) and λ_4 (Higgs quartic coupling). At trans-Planckian energies, gravitational fluctuations generate asymptotic safety in those couplings which do not become asymptotically free. At the Planck scale, gravitational fluctuations decouple and the renormalization group flow of the standard model couplings changes. Figure adapted from ref. 19 under a Creative Commons licence CC BY 4.0.

couplings. This makes gravity a 'non-renormalizable' theory, meaning that its predictivity is lost.

The non-renormalizability of gravity has prompted the development of various non-quantum field theory approaches to quantum gravity that postulate that the physics at small distances is radically different. Most prominent among them are perhaps string theory, which postulates that elementary particles are not pointlike, but are excitations of nonlocal objects (the strings), and loop quantum gravity, which postulates that space-time is fundamentally discrete, such that the quantization procedure is modified at small length scales.

Although perturbative renormalizability is a necessary prerequisite for asymptotic freedom, it is logically independent of asymptotic safety. In 1976, it was conjectured that asymptotic safety may describe the high-energy regime of quantum gravity[2]. It took twenty years before the development of functional renormalization group techniques made it possible to study the conjecture in four-dimensional gravity[3].

Following the breakthrough in functional renormalization group methods, asymptotic safety in gravity has become a well established field of study (see ref. 4 and references therein). Multiple studies have identified an interacting fixed point (reviewed in refs. 4–6), with non-zero couplings corresponding to the dimensionless counterparts of the Newton coupling, the cosmological constant and also various higher-curvature couplings that become important when describing gravity in conditions with large energy densities.

A key question is how many of these couplings are truly independent, and which can be determined as functions of the others. If just a finite number of the couplings are truly independent, then asymptotic safety can solve the predictivity problem raised by the result of perturbative renormalizability. Numerous papers provide evidence that asymptotically safe quantum gravity has three free parameters (see ref. 7, for example), which can be fixed by the low-energy value of the Newton coupling, the cosmological constant and a curvature-squared coupling, although more work is needed to definitely confirm this result[8].

The ratio of the cosmological constant to the square of the Planck mass is tiny. This is often considered a problem to be solved in quantum gravity because the Planck mass is expected to be the characteristic scale of all couplings in quantum gravity, including the cosmological constant. However, the observed ratio is in fact compatible with asymptotic safety, although the ratio is not predicted by the theory. In fact, different renormalization group trajectories starting from the microscopic fixed point lead to different possible values of the cosmological constant, one of which could describe our Universe (see Fig. 1a).

Although the field of asymptotically safe gravity has achieved encouraging results, important research questions remain. Resolving these is important to establishing the validity of this approach to quantum gravity.

One challenge in studying theories of gravity with the renormalization group is the proper treatment of space-time. A first difficulty arises because, in the renormalization group, a cutoff is imposed on the momentum associated with the Fourier modes of quantum fluctuations. This is easily possible for the spatial momentum, but not effective with the four-momentum, whose absolute value can be close to zero, even if the spatial momentum is large. In non-gravitational quantum field theory, this is dealt with by a Wick rotation of the time direction, which effectively converts time into just another spatial dimension. In quantum gravity, the Wick rotation is not well defined and calculations have to date mostly been done in purely spatial settings. Recently, first hints were obtained that asymptotic safety extends to space-time settings[9].

A second difficulty arises because the definition of a cutoff relies on the space-time metric, which determines the length and momentum scales. In quantum gravity, the space-time metric is subject to quantum fluctuations. This raises the question of which metric we

should choose to measure lengths and define the cutoff. In functional renormalization group techniques, an auxiliary background metric is chosen, and work is now being conducted to show that the results are independent of this choice[5,10].

Another open question concerns the dynamical properties of an asymptotically safe theory, which requires higher-curvature terms. These terms are an extension to the established Einstein action of general relativity. These interactions could produce new degrees of freedom with properties that are inconsistent with a theory of gravity, such as predictions of negative probability. However, such higher-order terms may also simply modify the behaviour of the quantized gravitational field, without introducing new degrees of freedom[11].

Given these challenges, it is critical to develop further methods that could corroborate asymptotic safety in gravity. Lattice techniques are an important example, including causal dynamical triangulations[12]; and functional renormalization group techniques have also been connected to other methods, such as minimal subtraction schemes[13].

The standard model and asymptotic safety

An asymptotically safe model of gravity could possibly provide an ultraviolet completion of the standard model. At energies above the Planck scale, quantum fluctuations of gravity could alter the renormalization group flow of the standard model to induce asymptotic safety, as found within the approximations made in refs. 14–16.

There is evidence that coupling to gravity improves the behaviour of the standard model, so that its divergent couplings instead flow from a fixed point at small length scales. However, the combination of gravitational and standard model quantum fluctuations can only balance out at specific values or within specific ranges of couplings. This balance generates structures in how the couplings flow at different energy scales (Fig. 1b)[17].

Gravitational fluctuations decouple from the standard model below the Planck scale. However, the structures imprinted on the standard model couplings at the Planck scale are mapped by the renormalization group flow to structures at much lower energies, close to those of the electroweak scale. This procedure achieves something that is rarely achieved in quantum gravity: a potential confrontation with experimental data. Examples include a prediction of the ratio of the Higgs mass to the electroweak scale[18], an upper bound on the Abelian gauge coupling[17] and the top Yukawa coupling[19] as well as on the ratio of the top to the bottom Yukawa couplings[20] (see also ref. 21 and references therein).

Comparisons between experimental data and calculations exploring asymptotic safety have been encouraging, because the predicted structures broadly match observations, albeit within large systematic uncertainties in the theory. A refinement of existing predictions by reducing the systematic uncertainties is an obvious next step for the field. It should then be possible to discover whether predictions for the standard model match experimental results. This would considerably strengthen the case for an asymptotically safe fixed point.

There is scope for additional predictions beyond the standard model, most importantly in the dark matter sector. These run along two distinct lines: first, not all dark-matter models may be amenable to an asymptotically safe ultraviolet completion with quantum gravity[22]. Those which are not amenable give rise to the first prediction of asymptotically safe gravity: that such forms of dark matter should not exist.

Second, dark matter models that do fit into the asymptotic-safety paradigm should be described by an interacting fixed point and should therefore be scale-symmetric at small length scales. As discussed above for the standard model, the renormalization group flow may fix or bound the values of couplings, resulting in corresponding predictions at energies accessible to experiment[23].

Making quantitatively precise predictions for physics beyond the standard model is an important new frontier of the field and may at the same time address a key challenge of particle physics beyond the standard model, namely the large number of models that exist. The asymptotic-safety paradigm may greatly reduce the number of theoretically viable models and make concrete predictions for experimental programmes such as dark matter searches.

Astrid Eichhorn
University of Southern Denmark, Odense, Denmark.
e-mail: eichhorn@sdu.dk

Published online: 9 November 2023

References

1. Litim, D. F. & Sannino, F. *J. High-Energy Phys.* **12**, 178 (2014).
2. Weinberg, S. Critical phenomena for field theorists. In *Understanding the Fundamental Constituents of Matter* (ed. Zichichi, A.) (Springer, 1978).
3. Reuter, M. *Phys. Rev. D* **57**, 971 (1998).
4. Bonanno, A. et al. *Front. Phys.* **8**, 269 (2020).
5. Pawlowski, J. M. & Reichert, M. *Front. Phys.* **8**, 527 (2021).
6. Saueressig, F. Preprint at https://arxiv.org/abs/2302.14152 (2023).
7. Falls, K., Ohta, N. & Percacci, R. *Phys. Lett. B* **810**, 135773 (2020).
8. Kluth, Y. & Litim, D. F. *J. High-Energy Phys.* **108**, 026005 (2023).
9. Fehre, J., Litim, D. F., Pawlowski, J. M. & Reichert, M. *Phys. Rev. Lett.* **130**, 081501 (2023).
10. Morris, T. R. *J. High-Energy Phys.* **11**, 160 (2016).
11. Draper, T., Knorr, B., Ripken, C. & Saueressig, F. *Phys. Rev. Lett.* **125**, 181301 (2020).
12. Loll, R. *Class. Quant. Grav.* **37**, 013002 (2020).
13. Baldazzi, A., Percacci, R. & Zambelli, L. *Phys. Rev. D* **103**, 076012 (2021).
14. Alkofer, R. et al. *Ann. Phys.* **421**, 168282 (2020).
15. Kowalska, K., Pramanick, S. & Sessolo, E. M. *J. High-Energy Phys.* **08**, 262 (2022).
16. Pastor-Gutiérrez, Á., Pawlowski, J. M. & Reichert, M. *SciPost Phys.* **15**, 105 (2023).
17. Eichhorn, A. & Versteegen, F. *J. High-Energy Phys.* **1801**, 30 (2018).
18. Shaposhnikov, M. & Wetterich, C. *Phys. Lett. B* **683**, 196–200 (2010).
19. Eichhorn, A. & Held, A. *Phys. Lett. B* **777**, 217–221 (2018).
20. Eichhorn, A. & Held, A. *Phys. Rev. Lett.* **121**, 151302 (2018).
21. Eichhorn, A. & M. Schiffer, M. Preprint at https://arxiv.org/abs/2212.07456 (2022).
22. de Brito, G. P., Eichhorn, A. & Lino dos Santos, R. R. *J. High-Energy Phys.* **6**, 13 (2022).
23. Reichert, M. & Smirnov, J. *Phys. Rev. D* **101**, 063015 (2020).

Competing interests

The author declares no competing interests.

https://doi.org/10.1038/s41567-023-02263-5

Supersymmetric renormalization group flow

Jaewon Song

Supersymmetric quantum field theories have special properties that make them easier to study. This Comment discusses how the constraints that supersymmetry places on renormalization group flows have been used to study strongly coupled field theories.

Ken Wilson's work on the renormalization group taught us that quantum field theories (QFT) are defined according to a characteristic energy or length scale. Typically, we start with a microscopic or ultraviolet description of a QFT. Renormalization group transformations then enable us to coarse-grain the theory. In the renormalization group procedure, the parameters of a theory 'flow' to produce an effective field theory that is valid at low energies or long distances. However, in most cases, renormalization group equations cannot be solved exactly. Therefore, results are often calculated by approximating the effect of small perturbations from solvable non-interacting theories.

In high-energy physics, so-called gauge theories play a particularly important role. These QFTs are characterized by an invariance under local symmetry transformations called a gauge transformation. For example, the fundamental strong, weak and electromagnetic forces are known to be described by gauge theories.

Some gauge theories, such as the one that describes the strong force, are asymptotically free. This means that at high energies, the fields become non-interacting. Conversely, in the low-energy or infrared limit, the interaction strength grows, and it becomes impossible to use perturbative approximation methods to determine the low-energy limit of the gauge theory.

A low-energy effective field theory must still satisfy the symmetry properties of the original microscopic theory. Ensuring that the symmetry is respected can constrain the dynamics of the strongly coupled effective theory.

Supersymmetry is a space-time symmetry between bosons and fermions. QFTs with supersymmetry enjoy constraints from holomorphy[1], which restrict the allowed terms in the effective field theory. This property made it possible to discover the phenomenon of infrared duality[2], meaning that distinct gauge theories describe the same physics at low energies (infrared) upon renormalization group flow. Moreover, developments in recent decades[3] have made it possible to analytically compute other observables, such as supersymmetric partition functions. These features allow us to test the infrared dynamics of supersymmetric theories with high precision.

In this Comment, we describe how supersymmetry helped us to better understand certain features of renormalization group flows. Supersymmetry enables some exact predictions, providing valuable insights into the physics of strongly coupled theories. We will focus mainly on the theories with four space-time dimensions.

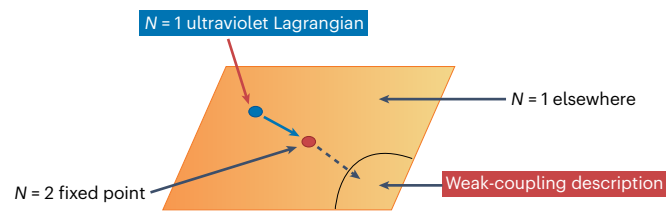

Fig. 1 | $N = 1$ theory flows in the infrared to a point on a conformal manifold of $N \geq 2$ theory. A non-Lagrangian $N = 2$ theory may be marginally deformed to an $N = 1$ theory with a weakly coupled dual description.

a-maximization

Let us consider a supersymmetric gauge theory in the ultraviolet. Generally the theory flows in the infrared to a fixed point at which the theory's parameters no longer change under renormalization group transformations. The fixed point can be trivial, with an energy gap between the ground state and the lowest-energy excitations. More interestingly, the theory can be gapless, with 'massless' degrees of freedom.

The constraints imposed by supersymmetry and invariance under renormalization mean that a massless theory must be a superconformal field theory (SCFT). The symmetry of a four-dimensional SCFT is described by the superconformal group $SU(2,2|N)$, where N denotes the number of independent supersymmetry transformations the SCFT has. In this section, we will restrict ourselves to the case of minimal supersymmetry, that is, $N = 1$.

The superconformal group $SU(2,2|1)$ contains a bosonic subgroup:

$$SO(4,2) \times U(1)_R \subset SU(2,2|1) \tag{1}$$

where $SO(4,2)$ is the conformal group containing the Poincaré group of space-time symmetries and the dilation generator associated with the scaling symmetry. The additional $U(1)_R$ group is the so-called R-symmetry, and this symmetry has a crucial role in the physics of SCFTs. For example, the scaling dimensions of certain operators in the SCFT are constrained by how they transform under the R-symmetry.

The stress tensor T of a conformal field theory represents the quantities that are conserved owing to translational invariance. In flat space, its trace is zero, but in curved space, it is non-vanishing with the following form:

$$\langle T^{\mu}_{\mu} \rangle = cW^2 - aE_4 \tag{2}$$

where W is the Weyl tensor and E_4 is the Euler density of the space-time manifold, both of which reflect the curvature. The coefficients a and c are called conformal anomalies or central charges. In particular, a is known to measure the degrees of freedom, since it decreases along the renormalization group flow[4,5]. Supersymmetry relates the conformal anomalies to the R-symmetry.

Treating a global symmetry as a gauge symmetry introduces an inconsistency, whose nature should be the same in both the weakly coupled ultraviolet microscopic theory and in the strongly interacting infrared effective theory. This inconsistency can be quantified and is called the 't Hooft anomaly. The 't Hooft anomaly is invariant under the renormalization group flow, which allows us to use perturbative ultraviolet calculations to draw conclusions about the low-energy strongly coupled theory. This procedure is called 't Hooft anomaly matching[6]. For an SCFT, the central charges are conveniently computable using the 't Hooft anomaly involving the R-symmetry, because the stress tensor and the R-symmetry currents are related by supersymmetry[7].

However, it is not always clear what form of R-symmetry will emerge at a strongly coupled infrared fixed point. SCFTs may have additional $U(1)$ symmetries, which can mix together to give multiplet candidates of R-symmetry. In these cases, the so-called a-maximization principle[8] applies: among the set of possible $U(1)_R$ symmetries, the superconformal R-symmetry should maximize the a-function.

Performing a-maximization reveals whether the fixed point of the renormalization group flow is an SCFT or not. First, to have an SCFT at the fixed point, there must be a real solution to the a-maximization problem. Second, the SCFT must be unitary, which puts constraints on the R-charges of the operators. If an R-symmetry that satisfies these conditions exists, the infrared fixed point can be given by a non-trivial SCFT. Moreover, knowing the exact R-charge enables us to compute detailed properties of the SCFT's operators.

Similar procedures to determine the superconformal R-symmetry exist in other space-time dimensions: c-extremization in two dimensions[9] and F-maximization (or Z-extremization) in three dimensions[10].

Landscape of SCFTs

Armed with a-maximization, we can explore the space of renormalization group flows and the superconformal fixed points obtainable from a supersymmetric gauge theory. Once an infrared fixed point has been identified, we can enumerate all the relevant operators.

Additional renormalization group flows are triggered by deforming the original gauge theory via gauge-invariant operators. If they are relevant, they create flows to different fixed points. Relevant operators also typically break some of the initial global symmetry, so to find new fixed points, we maximize the trial a-function over smaller symmetries, which results in a smaller value of a. This provides a supersymmetric demonstration of the so-called a-theorem, which is the conjecture that a should always decrease along renormalization group flows. However, this argument has caveats, as pointed out in the original paper[8].

The a-maximization procedure involves identifying all the $U(1)$ symmetries that can mix with the R-symmetry. However, such $U(1)$ symmetries can emerge along the renormalization group flow. Huge quantum corrections can cause some of the operators in the theory to decouple from the rest of the system and become free fields. These operators are then associated with an accidental $U(1)$ global symmetry that acts only on them. This process invalidates a-maximization calculations based solely on symmetries that are manifest in the initial ultraviolet theory. Another caveat is that the trial a-function is only a local maximum, so it may happen that a larger maximum exists in the restricted parameter space.

Considering the decoupled operators carefully, we also find a non-canonical type of gauge theory. For example, consider $SU(N)$ gauge theory with an adjoint chiral multiplet and a small number of fundamental chiral multiplets. This model flows to an interacting SCFT with a number of decoupled free fields. In the large-N limit, the scaling dimensions of gauge-invariant operators at the fixed point form a dense set[11], meaning that the gap in the operator dimensions scales as $1/N$, owing to the large quantum corrections to the dimensions of the fundamental fields that are non-perturbative. This 'dense spectrum' contrasts sharply with typical gauge theories whose scaling dimensions are discrete in the large-N limit.

By considering all possible relevant deformations, including the decouplings, we can extend the space of renormalization group flows in a given gauge theory. Even for a very simple field theory with $SU(2)$, the gauge group reveals a large set of non-trivial fixed points[12]. These non-trivial fixed points exhibit various non-trivial phenomena – emergent symmetry, decoupling of operators, and narrow distribution of the central charges a/c. Some of these fixed points even display supersymmetry enhancement, which we discuss below.

Supersymmetry-enhancing renormalization group flows

A QFT is usually introduced by writing a Lagrangian, which defines how the fields behave and interact. However, different theoretical techniques make it possible to identify QFTs that do not have a known (or useful) Lagrangian description. These 'non-Lagrangian' theories are difficult to study quantitatively because most methods are based on analysing the terms of a Lagrangian. In many cases it is not clear whether a QFT is genuinely non-Lagrangian or whether the Lagrangian is simply as yet undiscovered.

However, there exists an unambiguous notion of 'non-Lagrangian' SCFTs for which it is impossible to write a Lagrangian exhibiting the system's full amount of supersymmetry. For example, Argyres–Douglas theory[13] is an $N = 2$ SCFT, whose operator contents are inconsistent with the constraints imposed by an $N = 2$ supersymmetric Lagrangian. Similarly, there can be no purely $N = 3$ Lagrangian theory, and would-be $N = 3$ supersymmetric theories actually have larger $N = 4$ supersymmetry. Nevertheless, many non-Lagrangian $N = 2,3$ SCFTs have been constructed using techniques from string theory[14,15], and the list of such theories continues to grow[16,17].

Recently, it was found that certain non-Lagrangian theories can arise as renormalization group fixed points of Lagrangian theories with minimal ($N = 1$) supersymmetry. A superconformal fixed point of a particular $N = 1$ gauge theory was found to match the minimal Argyres–Douglas theory[13], which has $N = 2$ supersymmetry[18]. This discovery means we have an $N = 1$ ultraviolet Lagrangian description for a non-Lagrangian $N = 2$ SCFT. From this gauge theory description, we can compute important physical quantities, such as supersymmetric partition functions, that were previously unknown. Importantly, the ultraviolet fixed point contains 'dangerously irrelevant' operators whose coupling gets smaller during the initial renormalization group flow but that eventually becomes relevant and grows, drastically altering the fixed point of the flow.

Many other examples of such supersymmetry-enhancing renormalization group flows have now been discovered. More-general Argyres–Douglas theories can be constructed by choosing ultraviolet theories with different gauge groups. Lagrangians with singular, divergent couplings have also been used[19] to obtain $N = 2$ SCFTs with flavour symmetry described by the exceptional Lie groups[14].

Sometimes we can deform an $N = 2$ SCFT via $N = 1$ supersymmetry-preserving marginal operators. This means that we have a family of $N = 1$ SCFTs parameterized by the couplings to the marginal operators (or a continuous family of fixed points, called a conformal manifold)

connected to the $N = 2$ SCFT we started with. This sometimes allows us to find a dual $N = 1$ Lagrangian gauge theory, which can be smoothly connected to $N = 2$ or $N = 3$ non-Lagrangian theories[20–22]. (See Fig. 1 for an illustration.) As a final example, a particular $N = 1$ theory built out of non-Lagrangian SCFTs has been shown to flow in the infrared to a point in the conformal manifold of a maximally supersymmetric $N = 4$ super Yang–Mills theory[23], exhibiting the largest possible supersymmetry enhancement.

Conclusion

As we have discussed here, supersymmetry provides a valuable theoretical laboratory in which to explore the strong-coupling dynamics of QFTs. It has been a particularly fruitful approach for discovering non-trivial aspects of the renormalization group, such as infrared duality, conformal manifolds, decoupling of operators, emergent (super) symmetry, and so on. Some of these phenomena have then been found in non-supersymmetric models as well. We expect that supersymmetry will continue to provide insights and tools that help us to understand this universal language of physics – QFT – better.

Jaewon Song
Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon, Korea.
e-mail: jaewon.song@kaist.ac.kr

Published online: 9 November 2023

References

1. Seiberg, N. & Witten, E. *Nucl. Phys. B* **426**, 19–52 (1994); erratum **430**, 485–486 (1994).
2. Seiberg, N. *Nucl. Phys. B* **435**, 129–146 (1995).
3. Pestun, V. et al. *J. Phys. A* **50**, 440301 (2017).
4. Cardy, J. L. *Phys. Lett. B* **215**, 749–752 (1988).
5. Komargodski, Z. & Schwimmer, A. *J. High-Energy Phys.* **12**, 099 (2011).
6. 't Hooft, G. *NATO Sci. Ser. B* **59**, 135–157 (1980).
7. Anselmi, D., Freedman, D. Z., Grisaru, M. T. & Johansen, A. A. *Nucl. Phys. B* **526**, 543–571 (1998).
8. Intriligator, K. A. & Wecht, B. *Nucl. Phys. B* **667**, 183–200 (2003).
9. Benini, F. & Bobev, N. *Phys. Rev. Lett.* **110**, 061601 (2013).
10. Jafferis, D. L. *J. High-Energy Phys.* **5**, 159 (2012).
11. Agarwal, P., Lee, K.-H. & Song, J. *J. High-Energy Phys.* **10**, 49 (2021).
12. Maruyoshi, K., Nardoni, E. & Song, J. *Phys. Rev. Lett.* **122**, 121601 (2019).
13. Argyres, P. C. & Douglas, M. R. *Nucl. Phys. B* **448**, 93–126 (1995).
14. Minahan, J. A. & Nemeschansky, D. *Nucl. Phys. B* **482**, 142–152 (1996).
15. García-Etxebarria, I. & Regalado, D. *J. High-Energy Phys.* **3**, 083 (2016).
16. Gaiotto, D. *J. High-Energy Phys.* **8**, 034 (2012).
17. Xie, D. *J. High-Energy Phys.* **1**, 100 (2013).
18. Maruyoshi, K. & Song, J. *Phys. Rev. Lett.* **118**, 151602 (2017).
19. Gadde, A., Razamat, S. S. & Willett, B. *Phys. Rev. Lett.* **115**, 171604 (2015).
20. Razamat, S. S. & Zafrir, G. *J. High-Energy Phys.* **9**, 046 (2019).
21. Zafrir, G. *J. High-Energy Phys.* **12**, 098 (2020).
22. Zafrir, G. *J. High-Energy Phys.* **1**, 062 (2021).
23. Kang, M. J., Lawrie, C., Lee, K.-H. & Song, J. *Phys. Rev. Lett.* **130**, 231601 (2023).

Acknowledgements

The author thanks P. Agarwal, K. Intriligator, M. Kang, C. Lawrie, K.-H. Lee, K. Maruyoshi, E. Nardoni, S. Razamat, A. Sciarappa and Y. Tachikawa for collaborations and discussions. This work is supported by National Research Foundation of Korea (NRF) grant RS-2023-00208602 and also by the POSCO Science Fellowship of POSCO TJ Park Foundation.

Competing interests

The author declares no competing interests.

https://doi.org/10.1038/s41567-023-02234-w

Consequences of the renormalization group for perturbative quantum chromodynamics

Diogo Boito

The renormalization group is a key ingredient in methods of improving perturbative computations in particle physics. Here I briefly discuss its role in perturbative quantum chromodynamics and particularly the running of its coupling constant.

Ken Wilson's interpretation of renormalization had profound implications for particle physics and quantum chromodynamics (QCD) in particular. Wilson's picture assumes that quantum field theory has a high-energy, or ultraviolet, cutoff Λ. The physical origin of this cutoff depends on the theory in question. In some cases, there is a natural cutoff related to the inverse of a fundamental distance scale such as the atomic or the lattice spacing, which is usually the case in condensed matter physics. In particle physics such a cutoff could reflect a fundamental scale of the Universe.

Wilson's main observation was that physics at low energies must be independent of fluctuations that take place at high energies. One can obtain an effective theory that only describes the low-energy degrees of freedom by 'integrating out' the high-energy modes, removing fluctuations above a certain energy cutoff. In this procedure the coupling constants that describe the physical interactions have to change so that the effective theory describes the same low-energy physics as the full theory. Wilson's renormalization group equations describe the modifications to the couplings that are due to infinitesimal changes in the cutoff, which are realized in terms of differential equations.

Wilson's formulation of the renormalization group was fundamental for our understanding of quantum field theory and for the conceptual development of effective theories — it is one of the main tools in the present-day toolbox. In some practical applications to particle physics, however, this formulation of the renormalization group is not particularly convenient. We need to understand how quantities related to physical observables — not just coupling constants — behave under the renormalization group.

When working with the standard model of particle physics, most of the calculations of interest are done perturbatively using the celebrated Feynman diagrams, in which lines represent propagating particles and vertices represent interactions (see Fig. 1). Higher precision requires going beyond the leading order, introducing Feynman diagrams with loops, which require integration over the momenta flowing in the loops.

An immediate connection to the Wilsonian renormalization group equations could be made by introducing a cutoff Λ in the momentum integrals, removing the short-wavelength, high-energy fluctuations. A characteristic feature of loop calculations is the appearance of non-analytic logarithms that involve ratios of scales. Calculations with a sharp ultraviolet cutoff Λ produce terms such as $\log\left(\frac{\Lambda^2}{-p^2}\right)$, where p^2 is some energy-characteristic scale of the process. The ultraviolet divergence of the theory is regulated by the cutoff and would be manifest in the limit $\Lambda \to \infty$. Unfortunately, calculations with a sharp ultraviolet cutoff are very inconvenient because the cutoff can break gauge symmetries of the Lagrangian that are central to our understanding of gauge fields and interactions.

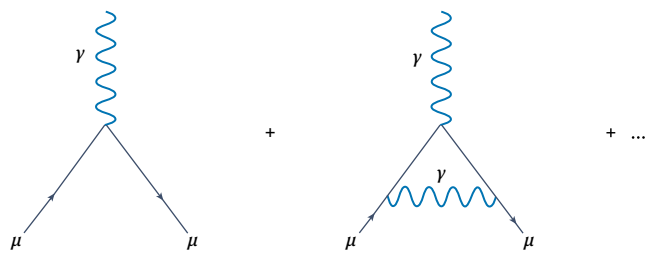

Fig. 1 | Dominant Feynman diagrams for the computation of muon magnetic moment. In such diagrams, solid lines correspond to muons μ and wave lines are photons γ. Vertices between lines represent interactions. The second diagram is a one-loop diagram first calculated[17] in 1948, and represents one of the first major successes of quantum electrodynamics: the prediction of the leading-order anomalous magnetic moments of leptons.

Dimensional regularization

The preferred alternative approach to isolating — or regularizing — loop-integral divergences while preserving gauge symmetries is known as dimensional regularization[1]. In dimensional regularization, momentum integrals are always performed assuming the cutoff $\Lambda \to \infty$. Divergences are avoided instead by performing the integrals in non-integer space-time dimensions $D = 4 - \epsilon$ and later expanding for $\epsilon \approx 0$. For example, in the one-loop diagram of Fig. 1, one needs to compute the following integral:

$$\mu^{(4-D)} \int \frac{d^D l}{(2\pi)^D} \frac{1}{l^2 - m^2} \frac{1}{(l-q)^2} \frac{1}{(k+l)^2 - m^2}, \qquad (1)$$

with m being the muon mass, k the momentum of the external photon and q the momentum of the incoming muon. One must introduce a renormalization scale μ to preserve the correct physical dimensionality of the predicted observables.

In dimensional regularization it is the scale μ that appears in the logarithms as $\log\left(\frac{\mu^2}{-p^2}\right)$ and, in order to avoid large logarithms, μ^2 must be of the order of the typical scale for the process, p^2. The logarithms always appear in combination with a term $2/\epsilon$ that encodes the ultraviolet divergences in the limit $\epsilon \to 0$, which is equivalent to $\Lambda \to \infty$. After the integrals have been properly regularized, the next step is to remove

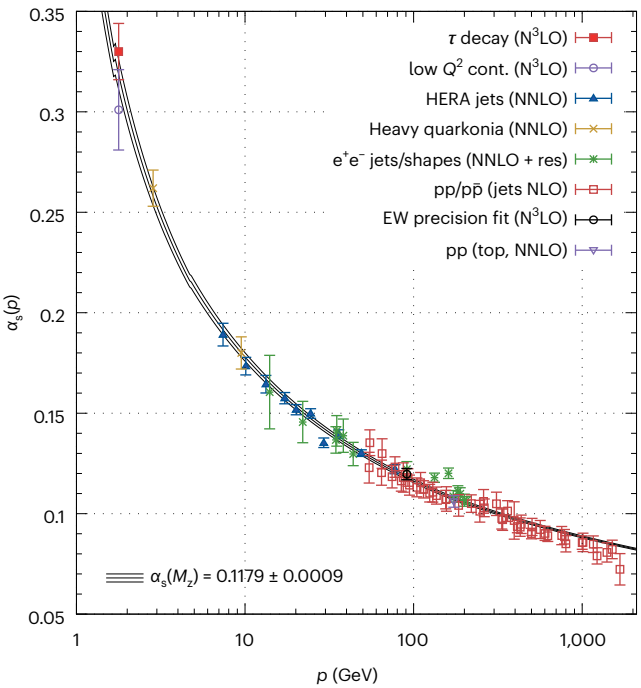

Fig. 2 | Running of the QCD coupling. Determinations of $\alpha_s(p)$ at different characteristic scales p. The solid lines represent the result obtained (within one uncertainty) from the world average of $\alpha_s(m_Z)$, solving the renormalization group equation (see equation (2)) with the β function calculated including terms up to five loops. The boundary condition is the average value of $\alpha_s(m_Z)$, obtained after evolving each individual determination to the Z boson mass scale. NLO, next-to-leading order; NNLO, next-to-next-to leading order; N^{3}LO, next-to-next-to-next-to leading order; NNLO + res, NNLO matched to a resumed calculation; EW, electroweak; cont., continuum. Adapted with permission from ref. 18, Oxford Univ. Press.

the $1/\epsilon$ divergences by suitable redefinitions of the Lagrangian parameters, leading to finite physical predictions at $D = 4$. The most frequently used renormalization scheme for these redefinitions is known as modified minimal subtraction. Here we assume that calculations are performed with dimensional regularization in the modified minimal subtraction scheme.

In the language of dimensional regularization, the dependence of coupling constants on the cutoff Λ translates into a dependence on the renormalization scale μ. However, the renormalization scale is not a physical quantity and physical predictions should not depend on it.

The independence of physical quantum field theory predictions with respect to the renormalization scale is the essence of the renormalization group in this context and it was first proved by Callan[2] and Symanzik[3]. It is realized as non-trivial cancellations between the μ-dependent logarithms that appear during loop computations and those that arise from the μ dependence of coupling strengths and other Lagrangian parameters.

Quantum chromodynamics

We focus now on the case of QCD, denoting by α_s the strong coupling, which can be thought of as the strong-interaction equivalent of the fine-structure constant. The function that governs the scale dependence of α_s is the QCD β function, which can be obtained perturbatively, as a series in increasing powers of α_s corresponding to Feynman diagrams with increasing numbers of interaction vertices. The renormalization group implies that

$$\frac{\mu}{2}\frac{d\alpha_s(\mu)}{d\mu} = \beta(\alpha_s) = \beta_0\alpha_s^2(\mu) + \beta_1\alpha_s^3(\mu) + \cdots \qquad (2)$$

because $\beta_0 < 0$, $\alpha_s(\mu)$ decreases with increasing μ. The interaction becomes less intense at high energies and we say that the theory becomes asymptotically free.

When the β function's first term, the one-loop coefficient β_0, was first calculated in 1973 (refs. 4,5), it provided an explanation for an effect known as Bjorken scaling, observed in the scattering of electrons on nucleons. Asymptotic freedom is one of the most important features of QCD and its discovery was the subject of the 2004 Nobel Prize in Physics.

Making progress in multi-loop calculations in realistic quantum field theories, and in QCD in particular, is notoriously difficult. At present, 50 years after the publication of the result for β_0, the QCD β function is now known at five loops[6–8]. Only recently in 2016 was the five-loop result, involving the computation of 1.5 million Feynman diagrams, finally published.

With the current level of precision, we are able to stringently constrain the scale dependence of $\alpha_s(\mu)$ predicted by the renormalization group equations. However, theory alone cannot give us the actual value of the coupling at any scale. The μ dependence of the coupling tells us that it is not a physical quantity, but it can be determined within a given renormalization convention (here the modified minimal subtraction scheme) from the careful comparison of state-of-the-art computations with experimental results for QCD observables or using first-principles lattice QCD simulations.

There are many determinations of α_s at different renormalization scales (Fig. 2). The agreement between the theory curve and the α_s determinations is excellent and spans three orders of magnitude in energy. The determination of α_s from the τ lepton decays (filled red square in Fig. 2) is particularly notable because it provides the most stringent test of the renormalization group equations at low energies. Asymptotic freedom is manifest in the fact that $\alpha_s(\mu)$ slowly goes to zero at high energies.

Another important feature is the divergence of the coupling at low energies. The point of divergence is known as the Landau pole and it occurs at a characteristic energy scale of the order of a few hundred megaelectronvolts (the precise value depends on the renormalization scheme). Because of this behaviour, perturbative QCD ceases to be valid for $p \lesssim 1.0$ GeV. The existence of the Landau pole may be connected to the non-perturbative phenomenon of quark confinement, which is still not fully understood.

Although physical quantities should be independent of μ, in practice all we have in perturbative QCD computations are truncated power series in $\alpha_s(\mu)$. The results obtained from these series retain a residual μ dependence. The key point is that this residual renormalization-scale dependence should become smaller at higher orders, because the perturbative series is systematically improvable by going to a higher number of loops.

For some time, this residual dependence was considered an unwanted feature that should be eliminated. A valuable theory effort was devoted to finding a unique, optimal, renormalization scale or scheme[9–11]. At present, however, the attitude of a significant part of the

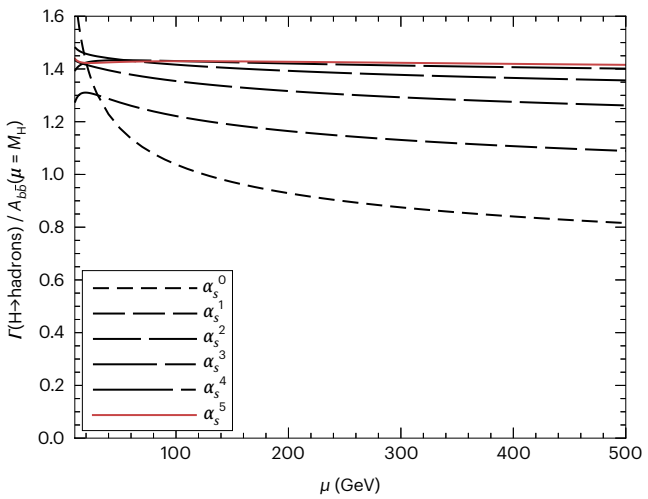

Fig. 3 | Renormalization scale dependence for the decay of the Higgs boson into bottom quarks and gluons. The decay rate Γ at an energy scale μ is normalized to the leading-order decay to bottom quarks ($A_{b\bar{b}}$) at the energy scale of the Higgs mass M_H. Different lines correspond to different orders of approximation $O(\alpha_s)$. The final result (red line) considers complete QCD corrections up to α_s^4 and the inclusion of α_s^5 terms in the decay to gluons) (with the mass of the bottom quark taken to be zero). Reproduced from ref. 12 under a Creative Commons licence CC BY 4.0.

theory community is that a residual scale dependence is essentially unavoidable. In fact, it provides a way to assess the theoretical error stemming from missing higher orders in perturbation theory. This is because, keeping n terms in the perturbative expansion, the residual scale dependence is always formally $O(\alpha_s^{n+1})$, a direct consequence of the renormalization group (equation (2)). The stability of the calculated value can be verified by how it changes when the renormalization scale is varied over a reasonable interval.

As a concrete example, Fig. 3 (reproduced from ref. 12) shows the result for the QCD corrections to the decay of the Higgs boson (H) into bottom quarks and gluons up to $O(\alpha_s^4)$. The improvement with respect to μ variation, order by order, is clear and the red line, which includes some contributions at $O(\alpha_s^5)$, shows an impressive μ independence, demonstrating that there is only a small error from missing higher-order terms.

Divergence of the perturbative expansion

In realistic quantum field theories, such as QCD, the perturbative series in powers of α_s is a divergent expansion, as discovered in the context of QED by Dyson in 1952 (ref. 13). However, the practical fact that truncated perturbative expansions in QED and QCD provide a very good description of experimental data very strongly suggests that the perturbative series are asymptotic series. This is the particular type of divergent expansions that provides a good approximation when truncated at intermediate orders but that, eventually, diverges.

Mathematically, the series is divergent because the order-n series coefficients, c_n, grow factorially with n. This means that, no matter how small the coupling may be, at high orders factorial growth takes over and it makes no sense to keep adding terms to the series. And — this is important — this factorial behaviour is not connected with the proliferation of Feynman diagrams. The origin of this divergence is, in fact, deeply rooted in the renormalization group and the renormalization procedure. The QCD β function and the evolution of α_s play a crucial role.

In Fig. 2, we see the manifestation of the Landau pole at low energies but, in dimensional regularization, the loop integrals are performed from zero to ∞ (in energy). Clearly, we cannot perform the loop integration in the low-energy region with impunity and the Landau pole must leave a trace in the final, finite, result. It leads, precisely, to perturbative coefficients that behave as $c_n \approx n!$. In fact, the coupling goes to zero rather slowly (logarithmically), as predicted by asymptotic freedom, and this also leaves a trace in the form of contributions to c_n that grow factorially but in this case with a sign alternation.

This type of factorial growth is better understood by applying an inverse Laplace transformation to the series, which in this context is called a Borel transformation. The factorials in the original function correspond to poles in the transformed variable. These poles are known as the renormalons of perturbation theory[14]. Effects due to renormalons are especially important at lower energy scales, because α_s is larger in this regime, such as, for example, in the description of hadronic τ decays (filled red square in Fig. 2). In this case, the complications due to renormalons led to an ambiguity in the renormalization scale setting that persisted for about 30 years, and was only very recently resolved[15,16].

Wilson's ideas were fundamental to our deep understanding of quantum field theory and, in particular, about the renormalization procedure. But, more than a crucial conceptual tool, they have led to the development of new systematic methods, such as effective field theories, and are used in practical applications to guide the improvement of perturbative computations, leading to high-precision results that are key to the testing of the standard model of particle physics.

Diogo Boito
Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil.
e-mail: boito@ifsc.usp.br

Published online: 9 November 2023

References

1. 't Hooft, G. & Veltman, M. J. G. *Nucl. Phys. B* **44**, 189 (1972).
2. Callan, C. G. Jr *Phys. Rev. D* **2**, 1541 (1970).
3. Symanzik, K. *Commun. Math. Phys.* **18**, 227 (1970).
4. Gross, D. J. & Wilczek, F. *Phys. Rev. Lett.* **30**, 1943 (1973).
5. Politzer, H. D. *Phys. Rev. Lett.* **30**, 1346 (1973).
6. Baikov, P. A., Chetyrkin, K. G. & Kühn, J. H. *Phys. Rev. Lett.* **118**, 082002 (2017).
7. Herzog, F., Ruijl, B., Ueda, T., Vermaseren, J. A. M. & Vogt, A. *J. High-Energy Phys.* **2**, 90 (2017).
8. Luthe, T., Maier, A., Marquard, P. & Schroder, Y. *J. High-Energy Phys.* **10**, 166 (2017).
9. Brodsky, S. J., Lepage, G. P. & Mackenzie, P. B. *Phys. Rev. D* **28**, 228 (1983).
10. Grunberg, G. *Phys. Rev. D* **29**, 2315 (1984).
11. Boito, D., Jamin, M. & Miravitllas, R. *Phys. Rev. Lett.* **117**, 152001 (2016).
12. Davies, J., Steinhauser, M. & Wellmann, D. *Nucl. Phys. B* **920**, 20 (2017).
13. Dyson, F. J. *Phys. Rev.* **85**, 631 (1952).
14. Beneke, M. *Phys. Rep.* **317**, 1–142 (1999).
15. Hoang, A. H. & Regner, C. *Phys. Rev. D* **105**, 096023 (2022).
16. Benitez-Rathgeb, M. A., Boito, D., Hoang, A. H. & Jamin, M. *J. High-Energy Phys.* **7**, 16 (2022).
17. Schwinger, J. S. *Phys. Rev.* **73**, 416 (1948).
18. Particle Data Group. et al. *Prog. Theor. Exp. Phys.* **8**, 083C01 (2022).

Acknowledgements

This work is supported by São Paulo Research Foundation (FAPESP) grant 2021/06756-6 and by CNPq grant 308979/2021-4.

Competing interests

The author declares no competing interests.

https://doi.org/10.1038/s41567-023-02196-z

The renormalization group for non-equilibrium systems

Yuhai Tu

Historically, most renormalization group studies have been performed for equilibrium systems. Here, I give a personal reflection on the unexpected outcome of studying non-equilibrium flocking using renormalization methods.

Ken Wilson's work on renormalization group theory developed from the study of critical phenomena in equilibrium systems (see ref. 1 for Wilson's review on the subject). Renormalization group methods had an immediate and lasting impact. However, in his Nobel lecture delivered in late 1982, Ken Wilson remarked: "In my view the extensive research that has already been carried out using the renormalization group and the ϵ expansion is only the beginning of the study of a much larger range of applications that will be discovered over the next twenty years (or perhaps the next century will be required)." In this Comment, I will describe my own experience developing one such application: the study of non-equilibrium stochastic dynamical systems.

The initial developments of the renormalization group theory by Leo Kadanoff and Ken Wilson were based on equilibrium systems. In these cases, it is possible to write a Hamiltonian that governs the statistical properties of the system.

The basic idea is to consider many-body systems at a coarse-grained level, averaging the microscopic degrees of freedom over a larger length scale. Kadanoff proposed[2] that a coarse-grained system can, after appropriate rescaling, be described by a Hamiltonian of the same form as the initial model but with a set of renormalized parameters. The transformation of parameters at a finer scale into those at a coarser scale constitutes the renormalization group.

Repeatedly applying this coarsening process produces a so-called flow of the parameters. This flow features fixed points — special values that do not change under renormalization. These parameters correspond to the critical point where the system obeys universal scaling laws. Systems with the same scaling laws are said to belong to the same universality class.

Another key idea in renormalization group theory is the existence of an upper critical dimension d_c. Above this dimension, the fixed point corresponding to mean field theory is stable, which means it correctly captures the critical behaviour. For lower dimensions, an approximation to the critical behaviour can be found by solving the renormalization group equations perturbatively[3] in orders of $\epsilon \equiv d_c - d$ to obtain the scaling exponents in leading orders of ϵ.

Subsequent work quickly extended the use of renormalization group to study dynamics of critical phenomena[4,5] (see ref. 6 for a comprehensive review). Although initially developed in equilibrium systems, it was clear that dynamic renormalization group can be used to study stochastic dynamics in non-equilibrium systems that do not have a Hamiltonian or other energy functionals. This is a very exciting perspective for renormalization group as most systems in nature are driven out of equilibrium by external forces, and yet they exhibit robust scaling behaviours similar to those observed in equilibrium critical phenomena.

One such example is the Kardar–Parisi–Zhang (KPZ) equation proposed to describe interface growth[7]. In the case of the KPZ equation, the renormalization group has been successfully used to understand its scaling behaviours[7–9]. On a personal level, studying the KPZ equation and its variance gave me the opportunity to learn and appreciate the beautiful intuition and powerful techniques behind renormalization[10,11]. Later, I worked with John Toner to develop a hydrodynamic theory of flocking where renormalization group methods played a critical role.

The most recognizable examples of flocking occur in bird flocks and fish schools (Fig. 1a,b). However, this type of collective motion can also arise at smaller length scales, for example in bacteria swarms (Fig. 1c), or in mixtures of motor proteins and microtubules.

Fig. 1 | Collective flocking in nature. Flocking behaviours span an enormous range of length scales from tens and hundreds of metres in **a**, fish schools and **b**, bird blocks, to micrometres in **c**, bacteria swarms.

Although flocking appears in many everyday contexts, it is only in the past 29 years or so that many of its universal features have been identified and understood. Much of this understanding has come from statistical physics ideas, applied in a series of studies performed by Tamas Vicsek and his collaborators, John Toner and myself.

John and I started working on flocking theory during a visit by Tamas Vicsek in late 1994 to the IBM Watson labs in Yorktown Heights, New York. Vicsek gave a seminar on a model for a population of active self-driven agents, referred to as boids[12]. Each boid moves with a constant speed in a direction derived from the average of its neighbours' directions of motion plus some noise. In their two-dimensional simulations of the model, Vicsek and his collaborators studied what happened when they lowered the noise level or increased the density of bodies. They observed a phase transition from a disordered phase in which the agents moved in random directions without long-range order to an ordered phase where they moved together coherently to form a flock.

At the end of his talk, Vicsek complained that he could not get his results published because the referee(s) insisted that the flocking state with long-range order is impossible in two dimensions as it violated the Mermin–Wagner theorem[13]. This theorem states that continuous symmetries cannot be broken in finite-temperature systems with short-range interactions in models with two or fewer dimensions.

This result seemingly implied that any finite noise will create fluctuations that destroy long-range order in a system with continuous symmetry and short-range interactions. The Vicsek model falls into this category, with the ordered state's single flocking direction breaking the model's rotational symmetry. It looked like spontaneous flocking, at least in two dimensions, was doomed.

John and I, both in the audience, asked a lot of questions during Vicsek's seminar. Puzzled by Vicsek's simulation results, we decided to work on the problem together immediately — perhaps the urgency came from the desire to save the boids from the seemingly unavoidable fate of moving apart in two dimensions. When John and I reconvened the next day, we already had the basic ingredients for a coarse-grained hydrodynamic theory. We kept in touch with Vicsek in the following months as we made more progress on the flocking theory, and Vicsek's paper was eventually published in 1995[14]. It is tempting to speculate that our work helped. The hydrodynamic theory for flocking that John and I developed is based on the dual roles of the boid's velocity field ($\mathbf{v}$). First, the model's interactions try to align a boid's velocity with those of its neighbours, which can be captured using a model known as vector ϕ^4 theory developed to describe magnetic materials. Second, the velocity vector also characterizes the physical motion of the boids, which transports the boid's body as well as the boid's velocity. This transport and flow can be captured by the Navier–Stokes equations.

Combining these two ingredients, we wrote down a dynamic field equation with both a convective term ($(\mathbf{v}\cdot\nabla)\mathbf{v}$) for transport and standard ϕ^4-type terms for the Heisenberg spin. Later, we extended the equation to include all possible relevant terms that preserve the rotational symmetry of both velocity and space without imposing the Galilean invariance, which is broken due to the unique reference frame, that is, the medium (air or water) within which boids move. This equation that describes the hydrodynamic properties of flocking is now called the Toner–Tu equation. It is worth emphasizing that due to the dual roles of the velocity field the flocking system is essentially out of equilibrium as there is no energy functional that can capture both the directional alignment and physical motion of the boids' dynamics.

Our analysis of the hydrodynamic theory using renormalization group techniques, was published[15] a few months after Vicsek's paper[14]. In the ordered phase where all boids move in the same direction, we could evaluate the stability of the system by computing the magnitude of fluctuations around the ordered state. The broken symmetry means that there are long-wavelength, low-energy degrees of freedom known as Goldstone modes that can be excited by noise sources.

Owing to the existence of the Goldstone modes, the variance of the velocity fluctuations around the ordered state contained a term that scales as $L^{2\chi}$ where L is the length scale of the system. We found that in the absence of nonlinear convective terms $\chi = 1 - d/2$, where d is the dimension of the system. Below three dimensions the fluctuations grow with the system size and the ordered state is unstable, in accordance with the Mermin–Wagner theorem.

However, the motion of boids in flocking systems makes the problem completely different from equilibrium models that are subject to the Mermin–Wagner theorem. In particular, we found the convective terms influence the model's renormalization group flow and that the exponent χ deviates from its value in the linear theory. In fact, by using certain invariant properties of our model that should apply for $d = 2$, we were able to obtain an exact expression for $\chi = \frac{3-2d}{5}$.

Our results show that $\chi = -1/5 < 0$ for $d = 2$ and the fluctuations around the long-range-ordered state remains finite in the thermodynamic limit — collective flocking behaviour is stable in two dimensions. Soon afterwards, we applied the hydrodynamic theory for flocking to study other important emergent phenomena, such as the sound waves[16] and the giant number fluctuations[17] of flocks. Most recently, the dynamics renormalization group approach has been used to study flocking behaviours in the presence of quenched disorder[18] and in natural swarms[19]. The flocking model has also become influential in areas outside of physics ranging from robotics[20] to traffic[21].

Beyond flocking, the versatility and power of the renormalization group-based approach have been demonstrated in different non-equilibrium systems such as active matter and living systems. For example, it has been used to develop[22] a coarse-graining approach for studying neural activities in a large network of neurons that reveals a quasi-universal scaling behaviour in neuron firing patterns in different parts of the brain[23]. In our own recent work, my colleagues and I constructed a state-space approach to understand the inverse power-law scaling of the energy dissipation rate across different scales in non-equilibrium reaction networks[24,25].

Non-equilibrium biological systems often exhibit unexpected behaviours that are drastically different from their equilibrium counterparts. Nevertheless, like equilibrium statistical physics models they exhibit collective behaviours across different scales. I believe that bringing fundamental ideas of coarse-graining and the renormalization group into the study of active and living systems is one of the most promising research directions in both physics and biology.

Yuhai Tu
IBM T. J. Watson Research Center, Yorktown Heights, NY, USA.
e-mail: yuhai@us.ibm.com

Published online: 9 November 2023

References

1. Wilson, K. G. *Rev. Mod. Phys.* **47**, 773–840 (1975).
2. Kadanoff, L. P. *Phys. Phys. Fiz.* **2**, 263–272 (1966).
3. Wilson, K. G. & Fisher, M. E. *Phys. Rev. Lett.* **28**, 240–243 (1972).
4. Halperin, B. I., Hohenberg, P. C. & Ma, S.-k *Phys. Rev. Lett.* **29**, 1548–1551 (1972).
5. Bausch, R., Janssen, H. K. & Wagner, H. *Z. Phys. B Condens. Matter* **24**, 113–127 (1976).
6. Hohenberg, P. C. & Halperin, B. I. *Rev. Mod. Phys.* **49**, 435–479 (1977).
7. Kardar, M., Parisi, G. & Zhang, Y.-C. *Phys. Rev. Lett.* **56**, 889–892 (1986).

8. Medina, E., Hwa, T., Kardar, M. & Zhang, Y.-C. *Phys. Rev. A* **39**, 3053–3075 (1989).
9. Canet, L., Chaté, H., Delamotte, B. & Wschebor, N. *Phys. Rev. Lett.* **104**, 150601 (2010).
10. Kessler, D. A., Levine, H. & Tu, Y. *Phys. Rev. A* **43**, 4551–4554 (1991).
11. Tu, Y. *Phys. Rev. Lett.* **73**, 3109–3112 (1994).
12. Reynolds, C. W. *SIGGRAPH Comput. Graph.* **21**, 25–34 (1987).
13. Mermin, N. D. & Wagner, H. *Phys. Rev. Lett.* **17**, 1133–1136 (1966).
14. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
15. Toner, J. & Tu, Y. *Phys. Rev. Lett.* **75**, 4326–4329 (1995).
16. Tu, Y., Toner, J. & Ulm, M. *Phys. Rev. Lett.* **80**, 4819–4822 (1998).
17. Toner, J. & Tu, Y. *Phys. Rev. E* **58**, 4828–4858 (1998).
18. Chen, L., Lee, C. F., Maitra, A. & Toner, J. *Phys. Rev. Lett.* **129**, 188004 (2022).
19. Cavagna, A. et al. *Nat. Phys.* **19**, 1043–1049 (2023).
20. Gazi, V. & Passino, K. *IEEE T. Automat. Contr.* **48**, 692–697 (2003).
21. Helbing, D. *Rev. Mod. Phys.* **73**, 1067–1141 (2001).
22. Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W. & Bialek, W. *Phys. Rev. Lett.* **123**, 178103 (2019).
23. Morales, G. B., di Santo, S. & Muñoz, M. A. *Proc. Natl Acad. Sci. USA* **120**, e2208998120 (2023).
24. Yu, Q., Zhang, D. & Tu, Y. *Phys. Rev. Lett.* **126**, 080601 (2021).
25. Yu, Q. & Tu, Y. *Phys. Rev. E* **105**, 044140 (2022).

Acknowledgements
The author would like to acknowledge insightful discussions with H. Chate.

Competing interests
The author declares no competing interests.

Additional information
Peer review information *Nature Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

https://doi.org/10.1038/s41567-023-02233-x

Rigorous renormalization group

Antti Kupiainen

The renormalization group evolved from ad hoc procedures to cope with divergences in perturbative calculations. This Comment summarizes efforts to develop a mathematically rigorous approach to renormalization group calculations.

Ever since quantum electrodynamics was first developed, the mathematical rigour of quantum field theory (QFT) has been called into question. Attempts to calculate quantities such as the magnetic moment of the electron perturbatively in the strength of interaction between electrons and photons had produced infinite answers.

The pioneers of quantum electrodynamics solved this problem with a systematic but ad hoc procedure called renormalization, whereby the infinities were packaged into a relation between the observable low-energy quantities – such as mass and charge – and the high-energy 'bare' quantities. The result was a well defined expansion of scattering amplitudes and the like in the former parameters. Unfortunately, the conceptual meaning of these manipulations was murky and, worse still, they did not offer a guide to understanding the weak, strong and gravitational interactions that seemed to be 'non-renormalizable' or too strongly interacting for these methods to work.

In mathematical physics, these difficulties were first addressed in the 1950s by asking what sort of mathematical object a quantum field theory (QFT) is. Quantum fields were viewed as operators $\Phi(x)$ acting in a Hilbert space H and depending on a space-time point x. Matrix elements $\langle\Omega|\Phi(x_1)\ldots\Phi(x_n)|\Omega\rangle$ with Ω the vacuum state in H define so-called Wightman functions of the space-time points x_i. Axioms were postulated describing the singularities and symmetries of the Wightman functions. Later, in the 1960s, a research programme known as constructive QFT began, with the aim of providing concrete examples satisfying the axioms from which to work towards a mathematical understanding of physically relevant QFTs.

An important input was provided by an observation going back to Julian Schwinger in 1958 that the Wightman functions have an analytic continuation to the Euclidean domain of imaginary time $x = (it, x)$ with $x \in R^3$ and t a real number. In that domain, the field operators commute and can be viewed as random (generalized) functions $\phi(x)$ on the Euclidean spacetime R^4. Furthermore, for physical theories described by a classical action functional $S(\phi)$ the probability distribution on field configurations ϕ has, formally, the density $e^{-\frac{1}{\hbar}S(\phi)}$. The Wightman functions become correlation functions of the random fields

$$\langle\phi(x_1)\ldots\phi(x_n)\rangle = \int \phi(x_1)\ldots\phi(x_n)e^{-\frac{1}{\hbar}S(\phi)}D\phi. \quad (1)$$

Therefore, the central problem of constructive QFT was to give a mathematically rigorous non-perturbative definition of the formal path integral in Eq. (1).

Divergences and scaling

In the 1970s this programme was carried out for the case of so-called superrenormalizable QFTs. The simplest example of these is the ϕ^4_d model, a theory of a scalar field $\phi : R^d \to R$ with classical action functional

$$S(\phi) = \int \left((\nabla\phi)^2 + m^2\phi^2 + \lambda\phi^4\right)dx, \quad (2)$$

where m is the field's mass and λ is the coupling constant for the field's interactions.

A natural approach is to give meaning to Eq. (1) by perturbing around the non-interacting case $\lambda = 0$. In that limit the field ϕ is described by a Gaussian distribution with covariance given by the Green function of the operator $(-\Delta + m^2)$:

$$\langle\phi(x)\phi(y)\rangle = (-\Delta + m^2)^{-1}(x, y).$$

Since this Green function diverges as $x \to y$, this field is not a function but rather a generalized function known as a distribution. This singular behaviour makes the definition of Eq. (1) as a perturbation of the $\lambda = 0$ case problematic because one needs to make sense of $\phi(x)^4$, with four fields evaluated at the same point. And indeed, a formal perturbation theory for Eq. (1) in powers of λ leads to divergent expressions, which is known as the ultraviolet problem. Likewise, for $m^2 = 0$ the field is strongly correlated at long distances with the Green function:

$$\langle\phi(x)\phi(y)\rangle \approx |x - y|^{2-d}.$$

In this case, perturbation theory gives rise to expressions that grow with the system size, dubbed infrared divergences.

One way to resolve these issues is to regularize the theory by introducing cutoffs. For example, an ultraviolet cutoff can be realized by replacing the continuum R^d by a lattice $(\epsilon Z)^d$ with spacing ϵ, and an infrared cutoff by restricting the interaction to a finite box. This turns the problem into a classical statistical mechanics model of a continuous spin on a lattice.

For the ultraviolet problem one can then try to find 'bare' parameters m_ϵ and λ_ϵ as functions of the cutoff ϵ so that the correlation functions have a limit as $\epsilon \to 0$. The superrenormalizable case corresponds to $d < 4$. Then the perturbative renormalization theory predicts that only a simple mass renormalization is needed to cancel the ultraviolet divergences. One simply replaces the physical mass m^2 in the action functional by the bare mass $m^2 + \delta m^2(\epsilon)$, with $\delta m^2(\epsilon) = a\lambda\log\epsilon$ in $d = 2$ and $\delta m^2(\epsilon) = b\lambda/\epsilon + c\lambda^2\log\epsilon$ in $d = 3$ with explicit coefficients a, b and c. However, for $d = 4$ an infinite series with divergent coefficients (known as counterterms) is needed for both the mass m and the coupling constant λ.

This problem of a divergent relation between the bare and the physical parameters is precisely what Ken Wilson addressed with his renormalization group method, following work by Leo Kadanoff.

The Kadanoff–Wilson renormalization group approach to QFT and critical phenomena is to view the problem using effective actions S_l that describe the physics at a particular spatial scale l.

The action S_l is defined by coarse-graining the field ϕ to a field ϕ_l, which has a short distance cutoff l. For instance, in the block spin scheme one takes $\phi_l(x)$ to be defined on the lattice $x \in (lZ)^d$ as an average of ϕ in a cube of side l centred at x. Then e^{-S_l} is defined as the probability density ϕ_l inherited from the probability distribution of ϕ. The renormalization group flow is then the map $l \to S_l$ that describes how the physics changes with scale.

In this framework, the ultraviolet problem can be posed as follows. Let $S^{(\epsilon)}$ be of the same form as Eq. (2) but defined on the lattice $(\epsilon Z)^d$ and with ϵ dependent bare parameters m_ϵ and λ_ϵ. We coarse grain $S^{(\epsilon)}$ to obtain $S_l^{(\epsilon)}$ and determine the 'bare' parameters so that $S_l^{(\epsilon)}$ has a limit as ϵ for all $l > 0$. For the infrared problem the ϵ is fixed, say to $\epsilon = 1$, and one inquires about the behaviour of S_l as $l \to \infty$.

A scaling argument reveals the role of the dimension d in this problem. For the case $m^2 = \lambda = 0$ in Eq. (2), the model is scale invariant. Indeed, the scaled field $l^{\frac{d-2}{2}}\phi(lx)$ has the same statistics as the field ϕ. Substituting this form into Eq. (2) leads us to expect that the scaled field is distributed with parameters $(lm, l^{4-d}\lambda)$. Then, if $d < 4$, the parameter λ is suppressed at small spatial scales $l \to 0$. If $d > 4$, the parameter λ is suppressed at large scales, $l \to \infty$. Thus the ultraviolet, short-range behaviour should be close to the non-interacting Gaussian limit if $d < 4$, and if $d > 4$ the infrared behaviour for the massless theory should be Gaussian. Although more detailed calculations give corrections to this simple scaling argument, the conclusions about the spatial dimension hold nevertheless.

Block spins

In the block spin scheme it is convenient to rescale distances so that all the effective actions are defined on a fixed lattice, such as the unit lattice Z^d. In this approach it is traditional to view the rescaled effective actions S_l as Hamiltonians H_l describing statistical mechanical models of continuous spins $\varphi(x) \in R$ with $x \in Z^d$. The coarse graining is obtained by fixing an integer $L > 1$, defining block spins as averages of the spins in lattice cubes of side L and rescaling back to the unit lattice. Hence, the renormalization group proceeds in discrete steps $H_l \to H_{Ll}$ and takes the form

$$e^{-H_{Ll}(\varphi')} = \int e^{-H_l(\varphi)}\delta(\varphi' - C\varphi)D\varphi \tag{3}$$

where $C\varphi$ is the coarse-graining and rescaling operation for block spins. The renormalization group flow of H_l is thus obtained by the iteration of a fixed map R acting on Hamiltonians: $H_{Ll} = RH_l$.

Let us apply the coarse-graining map Eq. (3) to a local Hamiltonian H of the form

$$H(\varphi) = \sum_{x\in\Lambda}[(\nabla\varphi(x))^2 + \mu\varphi(x)^2 + g\varphi(x)^4] \tag{4}$$

where μ and g are coupling constants and $\Lambda \subset Z^d$ is a finite box of lattice sites. The crucial observation going back to Wilson is that the path integral Eq. (3) is non-critical because it is already defined with a ultraviolet cutoff 1 and the coarse graining provides an infrared cutoff L.

Furthermore, the local nature of block spin transformation means that RH should also be approximately local, allowing for the iteration of the map R. If we start with small g we can attempt to expand the exponential in Eq. (3) in powers of g and evaluate each term using Gaussian integration. The result will be of the form

$$(RH)(\varphi) = \sum_{m=0}^{\infty}\sum_{x_1,\dots,x_m\in\Lambda'} K_m(x_1,\dots,x_m)\varphi(x_1)\dots\varphi(x_m) \tag{5}$$

where $\Lambda' = L^{-1}\Lambda$ is the rescaled box and $K_m(x_1,\dots,x_m)$ are many-body interactions, each of which is given by a formal perturbation series in powers of g and coefficients having rapid decay in the separations $x_i - x_j$. However, this formal perturbation series diverges with the nth Taylor coefficient growing as $n!$. The reason for this divergence can be traced back to the faster-than-quadratic growth of the φ interaction as $\varphi \to \infty$. This divergence is one of the main difficulties to overcome.

The rigorous approach[1] views Eq. (3) as a problem of classical statistical mechanics in the high-temperature regime and applies high-temperature cluster expansion methods to it. This results in a representation of the Gibbs factor as a gas of polymers

$$e^{-(RH)(\varphi')} = \sum_{k=0}^{\infty}\sum_{\{X_1,\dots,X_k\}}\prod_{i=1}^{k}\rho_{X_i}(\varphi') \tag{6}$$

where $X_i \subset \Lambda'$ are disjoint sets of lattice points called polymers, and $\rho_{X_i}(\varphi')$ depends on the spins $\varphi'(x)$ only for $x \in X_i$. The weight ρ_X of a polymer decays rapidly with the size of X. In a region of space where the field φ is not too large, it turns out that one may exponentiate Eq. (6) and recover an expansion of the form of Eq. (5), which turns out to be a convergent series! In other regions where φ is large, ρ_X is approximately given by $\exp[-cg\sum\varphi'(x)^4]$, reflecting the fact that large values of the field are improbable. These terms correspond to non-perturbative contributions that are known as instantons in the physics literature.

To iterate this renormalization group map one then needs to consider all the so-called relevant and marginal terms of the expansion (5) that do not contract under the linearization of the map $H \to RH$ around the Gaussian fixed point. For the model in Eq. (4) and $d = 4$ they turn out to be the same local terms already occurring in the Hamiltonian of Eq. (4). All the other terms in the expansion (5) form an infinite dimensional space of irrelevant perturbations that contract under the linearization.

Furthermore, the marginal interaction parameter g turns out to contract owing to contributions at second order in g. Thus, for the infrared problem, in order to construct the critical theory with infinite correlation length one has to fine-tune the relevant parameter μ of H as a function of g so that R^nH tends to the Gaussian fixed point of R as $n \to \infty$. Finally, all this will only work provided one can show that large values of the coarse-grained field remain improbable for all n. This procedure was carried out in ref. 1 where the Gaussian (mean field) behaviour

$$\langle\varphi(x)\varphi(y)\rangle \approx Z|x-y|^{-2}, |x-y| \to \infty. \tag{7}$$

was proved for the critical theory. A similar result was also proved in ref. 2 using a different method.

Implications for QFTs and outlook

As for the original goal of rigorously constructing a renormalizable QFT, the above analysis shows that this is not possible for the ϕ^4 model in four dimensions if we want to stay in the perturbative, small-g regime during the renormalization group process. Reaching a lengthscale l for a theory with ultraviolet cutoff ϵ requires the

renormalization group map to be applied $\log_L(l/\epsilon)$ times. Because g gets smaller with each renormalization group iteration, the procedure must begin with a very large 'bare' value if ϵ is small enough. Hence, it will be beyond the reach of perturbative analysis. Therefore, under this perturbative condition the $\epsilon \to 0$ limit is necessarily Gaussian. The nonperturbative result where g is allowed to be arbitrary has recently been established[3], in which it was proved that the ϵ limit is always Gaussian no matter how we choose the bare parameters as a function of the cutoff.

However, this raises the question of how to reconcile the existence of the renormalized perturbation series in the interaction g with the fact that the model must become non-interacting as the ultraviolet cutoff vanishes. The same issue arises for quantum electrodynamics and the electroweak part of the standard model, both of which are asymptotically free in the infrared but not in the ultraviolet.

The resolution of this dilemma is to consider S_l for different values of l as effective field theories for $l \geq \epsilon$ where the value of ϵ can be taken to be very small. Then the renormalized perturbation series is an asymptotic series for the correlation functions with this cutoff with great accuracy. For the scalar theory in Eq. (2) even more is true: one can define the theory by analytic continuation for negative g and then take the cutoff ϵ to zero. The renormalized perturbation series is then exactly an asymptotic series. However, the resulting QFT is not unitary and is therefore unphysical.

The prime examples of ultraviolet asymptotically free theories in four dimensions are non-abelian Yang–Mills theories. Some progress in their rigorous construction using renormalization group methods in finite volume has been achieved[4]. For these theories the infrared behaviour is non-perturbative: that is, the effective coupling constant increases with scale. In two dimensions the Gross–Neveu model of a self-interacting Dirac fermion is ultraviolet asymptotically free and it has been rigorously constructed in refs. 5,6. In the fermionic setup, interesting condensed matter problems have been rigorously studied using renormalization group methods[7].

The polymer expansion approach to the renormalization group method used in Eq. (6) has been widely applied[8] and extended to the study of first-order phase transitions and to disordered systems[9]. However, it has remained restricted to situations that are close to Gaussian; a more nonperturbative formalism is still missing. This situation has its parallel in numerical approaches to renormalization group methods where it has been very hard to go beyond approximative renormalization group schemes in which the renormalization group map is truncated in order to keep the effective actions local. Adding nonlocal corrections to such schemes has not resulted in numerical improvements. An alternative approach to renormalization group methods using tensor networks to produce effective actions that stay strictly local seems more promising from this point of view. The first steps to its rigorous analysis have been taken[10].

Antti Kupiainen
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.
e-mail: antti.kupiainen@helsinki.fi

Published online: 9 November 2023

References

1. Gawedzki, K. & Kupiainen, A. *Commun. Math. Phys.* **99**, 197–252 (1985).
2. Feldman, J., Magnen, J., Rivasseau, V. & Seneor, R. *Commun. Math. Phys.* **109**, 437–480 (1987).
3. Aizenman, M. & Duminil-Copin, H. *Ann. Math.* **194**, 163–235 (2021).
4. Balaban, T. *Commun. Math. Phys.* **122**, 355–392 (1989).
5. Feldman, J., Magnen, J., Rivasseau, V. & Seneor, R. *Commun. Math. Phys.* **103**, 67–103 (1986).
6. Gawedzki, K. & Kupiainen, A. *Commun. Math. Phys.* **102**, 1–30 (1985).
7. Giuliani, A., Mastropietro, V. & Porta, M. *Commun. Math. Phys.* **349**, 1107–1161 (2017).
8. Bauerschmidt, R., Brydges, D. C. & Slade, G. *Introduction to a Renormalisation Group Method* Lecture Notes in Mathematics 2242 (Springer, 2019).
9. Bricmont, J. & Kupiainen, A. *Commun. Math. Phys.* **116**, 539–572 (1988).
10. Kennedy, T. & Rychkov, S. *J. Stat. Phys.* **187**, 33 (2022).

Acknowledgements
A.K. was supported by the Academy of Finland and European Research Council Advanced Grant 741487.

Competing interests
The author declares no competing interests.