Note on renormalization group procedures

Stanisław D. Głazek
Institute of Theoretical Physics, Faculty of Physics, University of Warsaw
(Dated: 19 March 2014)

Concepts of renormalization group procedure are outlined here for the purpose of explaining a transition from the well-known integrating out of high energy components to the similarity procedure and eventually to the renormalization group procedure for effective particles, or RGPEP. This note was originally prepared as an item for the Wikipedia page on light-front quantization.

PACS numbers: 11.10.Gh, 11.10.Hi, 11.30.Cp

A. Renormalization group

Renormalization concepts, especially the *renormalization group* methods in quantum theories and statistical mechanics, have a long history and a very broad scope. The concepts of renormalization that appear useful in theories quantized in the front form of dynamics are essentially of two types, as in other areas of theoretical physics. The two types of concepts are associated with two types of theoretical tasks involved in applications of a theory. One task is to calculate observables (values of operationally defined quantities) in a theory that is unambiguously defined. The other task is to define a theory unambiguously. This is explained below.

Since the front form of dynamics aims at explaining hadrons as bound states of quarks and gluons and the binding mechanism is not describable using perturbation theory, the definition of a theory needed in this case cannot be limited to perturbative expansions. For example, it is not sufficient to construct a theory using regularization of loop integrals order-by-order and correspondingly redefining the masses, coupling constants, and field normalization constants also order-by-order. In other words, one needs to design the Minkowski space-time formulation of a relativistic theory that is not based on any *a priori* perturbative scheme. The front form of Hamiltonian dynamics is perceived by many researchers as most suitable framework for this purpose among the known options [1].

The desired definition of a relativistic theory involves calculations of as many observables as one must use in order to fix all the parameters that appear in the theory. The relationship between the parameters and observables may depend on the number of degrees of freedom that are included in the theory.

For example, consider virtual particles in a candidate formulation of the theory. Formally, special relativity requires that the range of momenta of the particles is infinite because one can change a momentum of a particle by an arbitrary amount through a change of frame of reference. If the formulation is not to distinguish any inertial frame of reference, the particles must be allowed to carry any value of momentum. Since the quantum field modes corresponding to particles with different momenta form different degrees of freedom, the requirement of including infinitely many values of momentum means that one requires the theory to involve infinitely many degrees of freedom. But for mathematical reasons, being forced to use computers for sufficiently precise calculations, one has to work with a finite number of degrees of freedom. One must limit the momentum range by some cutoff.

Setting up a theory with a finite cutoff for mathematical reasons, one hopes that the cutoff can be made sufficiently large to avoid its appearance in observables of physical interest. But in local quantum field theories that are of interest in hadronic physics the situation is not that simple. Namely, particles of different momenta are coupled through the dynamics in a nontrivial way and the calculations aiming at predicting observables yield results that depend on the cutoffs. Moreover, they do so in a diverging fashion.

There may be more cutoff parameters than just for momentum. For example, one may assume that the volume of space is limited, which would interfere with translation invariance of a theory, or assume that the number of virtual particles is limited, which would interfere with the assumption that every virtual particle may split into more virtual particles. All such restrictions lead to a set of cutoffs that becomes a part of a definition of a theory.

Consequently, every result of a calculation for any observable $X_{\rm observable}(\mu)$ characterized by its physical scale μ has the form of a function of the set of parameters of the theory, p, the set of cutoffs, say Λ, and the scale μ. Thus, the results take the form

$$X_{\rm observable}(\mu) = X_{\rm theory}(p, \Lambda, \mu) \, . \tag{1}$$

However, experiments provide values of observables that characterize natural processes irrespective of the cutoffs in a theory used to explain them. If the cutoffs do not describe properties of nature and are introduced merely for making a theory computable, one needs to understand how the dependence on Λ may drop out from $X_{\rm theory}(p, \Lambda, \mu)$. The cutoffs may also reflect some natural features of a physical system at hand, such as in the model case of an ultraviolet cutoff on the wave vectors of sound waves in a crystal due to the spacing of atoms in the crystal lattice. The natural cutoffs may be of enormous size in comparison to the scale μ. Then, one faces the question of how it happens in the theory that its results for observables at scale μ are not also of the enormous size of the cutoff and, if they are not, then how they depend on the scale μ.

The two types of concepts of renormalization mentioned above are associated with the following two questions:

1. How should the parameters p depend on the cutoffs Λ so that all observables $X(p, \Lambda, \mu)$ of physical interest do not depend on Λ, including the case where one removes the cutoffs by sending them formally to infinity?

2. What is the required set of parameters p?

The renormalization group concept associated with the first question [3] predates the concept associated with the second question [4], see [5]. Certainly, if one were in possession of a good answer to the second question, the first question could also be answered. In the absence of a good answer to the second question, one may wonder why any specific choice of parameters and their cutoff dependence could secure cutoff independence of all observables $X(p, \Lambda, \mu)$ with finite scales μ.

The renormalization group concept associated with the first question above relies on the circumstance that some finite set $p(\Lambda)$ yields the desired result,

$$X_{\rm observable}(\mu) = \lim_{\Lambda \to \infty} X_{\rm theory}[p(\Lambda), \Lambda, \mu] \, . \tag{2}$$

In this way of thinking, one can expect that in a theory with n parameters a calculation of n observables at some scale μ is sufficient to fix all parameters as functions of Λ. So, one may hope that there exists a collection of n effective parameters at scale μ, corresponding to n observables at scale μ, that are sufficient to parametrize the theory in such a way that predictions expressed in terms of these parameters are free from dependence on Λ. Since the scale μ is arbitrary, a whole family of such n-parameter sets labeled by μ should exist and every member of that family corresponds to the same physics. Moving from one such family to another by changing one value of μ to another is described as action of "the renormalization group." The word group is justified because the group axioms are satisfied; two such changes form another such change, one can invert a change, etc.

The question remains, however, why fixing the cutoff dependence of n parameters p on Λ using n conditions that n selected observables do not depend on Λ is good enough to make all observables in the physical range of μ not depend on Λ. In some theories such miracle may happen but in others it may not. The ones where it happens are called renormalizable, because one can normalize the parameters properly to obtain cutoff independent results.

Typically, the set $p(\Lambda)$ is established using perturbative calculations that are combined with models for description of non-perturbative effects. For example, perturbative QCD diagrams for quarks and gluons are combined with the parton models for description of binding of quarks and gluons into hadrons. The set of parameters $p(\Lambda)$ includes cutoff dependent masses, charges and field normalization constants. The predictive power of a theory set up this way relies on the circumstance that the required set of parameters is relatively small. The regularization is designed order-by-order so that as many formal symmetries as possible of a local theory are preserved and employed in calculations, as in the dimensional regularization of Feynman diagrams. The claim that the set of parameters $p(\Lambda)$ leads to finite, cutoff independent limits for all observables is qualified by the need to use some form of perturbation theory and inclusion of model assumptions concerning bound states.

The renormalization group concept associated with the second question above is conceived to explain how it may be so that the concept of renormalization group associated with the first question can make sense, instead of being at best a successful recipe to deal with divergences in perturbative calculations [6]. Namely, to answer the second question, one designs a calculation (see below) that identifies the required set of parameters to define the theory, the starting point being some specific initial assumption, such as some local Lagrangian density which is a function of field variables and needs to be modified by including all the required parameters. Once the required set of parameters is known, one can establish a set of observables that are sufficient to define the cutoff dependence of the required set. The observables can have any finite scale μ and one can use any scale μ to define the parameters $p(\Lambda)$, up to their finite parts that must be fitted to experiment, including features such as the observed symmetries.

Thus, not only the possibility that a renormalization group of the first type may exist can be understood, but also the alternative situations are found where the set of required cutoff dependent parameters does not have to be finite. Predictive power of latter theories results from known relationships among the required parameters and options to establish all the relevant ones [7].

The renormalization group concept of the second kind is associated with the nature of the mathematical computation used to discover the set of parameters p. In its essence, the calculation starts with some specific form of a theory with cutoff Λ and derives a corresponding theory with a smaller cutoff, in the sense of more restrictive, say $\Lambda/2$. After re-parameterization using the cutoff as a unit, one obtains a new theory of similar type but with new terms. This means that the starting theory with cutoff Λ should also contain such new terms for its form to be consistent with the presence of a cutoff. Eventually, one can find a set of terms that reproduces itself up to changes in the coefficients of the required terms. These coefficients evolve with the number of steps one makes, in each and every step reducing the cutoff by factor of 2 and rescaling variables. One could use other factors than 2, but 2 is convenient.

In summary, one obtains a trajectory of a point in a space of dimension equal to the number of required parameters and motion along the trajectory is described by transformations that form new kind of a group. Different initial points might lead to different trajectories. But if the steps are self-similar and reduce to a multiple action of one and the same transformation, say T, one may describe what happens in terms of the features of T, called the renormalization group transformation. The transformation T may transform points in the parameter space making some of the parameters decrease, some grow, and some stay unchanged. It may have fixed points, limit cycles, or even lead to chaotic motion.

Suppose that T has a fixed point. If one starts the procedure at this point, an infinitely long sequence of reductions of the cutoff by factors of 2 changes nothing in the structure of the theory, except the scale of its cutoff. This means that the initial cutoff can be arbitrarily large. Such theory may possess the symmetries of special relativity, since there is no price to pay for extending the cutoff as required when one wishes to make the Lorentz transformation that yields momenta which exceed the cutoff.

Both concepts of the renormalization group can be considered in quantum theories constructed using the front form of dynamics. The first concept allows one to play with a small set of parameters and seek consistency, which is a useful strategy in perturbation theory if one knows from other approaches what to expect. In particular, one may study new perturbative features that appear in the front form of dynamics, since it differs from the instant form. The main difference is that the front variables x^- (or p^+) are considerably different from the transverse variables $x^\perp$ (or $p^\perp$), so that there is no simple rotational symmetry among them.

One can also study sufficiently simplified models for which computers can be used to carry out calculations and see if a procedure suggested by perturbation theory may work beyond it. The second concept allows one to address the issue of defining a relativistic theory *ab initio* without limiting the definition to perturbative expansions. This option is particularly relevant to the issue of describing bound states in QCD. However, to address this issue one needs to overcome certain difficulties that the renormalization group procedures based on the idea of reduction of cutoffs are not capable of easily resolving. To avoid the difficulties, one can employ the similarity renormalization group procedure. Both the difficulties and similarity are explained in the next section.

B. Similarity transformations

A glimpse of difficulties of the procedure of reducing a cutoff Λ to cutoff $\Lambda/2$ in the front form of Hamiltonian dynamics of strong interactions can be gained by considering the eigenvalue problem for the Hamiltonian H,

$$H\psi = E\psi, \tag{3}$$

where $H = H_0 + H_I$, H_0 has a known spectrum and H_I describes the interactions. Let us assume that the eigenstate ψ can be written as a superposition of eigenstates of H_0 and let us introduce two projection operators, P and Q, such that P projects on eigenstates of H_0 with eigenvalues smaller than $\Lambda/2$ and Q projects on eigenstates of H_0 with eigenvalues between $\Lambda/2$ and Λ. The result of projecting the eigenvalue problem for H using P and Q is a set of two coupled equations

$$H_0 Q\psi + QH_I Q\psi + QH_I P\psi = EQ\psi\,, \tag{4}$$

$$H_0 P\psi + PH_I Q\psi + PH_I P\psi = EP\psi\,. \tag{5}$$

The first equation can be used to evaluate $Q\psi$ in terms of $P\psi$,

$$Q\psi = \frac{1}{E - H_0 - QH_I Q}\, QH_I P\psi\,. \tag{6}$$

This expression allows one to write an equation for $P\psi$ in the form

$$H_{\rm eff} P\psi = EP\psi\,, \tag{7}$$

where

$$H_{\rm eff} = H_0 + PH_I P + PH_I Q\, \frac{1}{E - H_0 - QH_I Q}\, QH_I P\,. \tag{8}$$

The equation for $P\psi$ appears to resemble an eigenvalue problem for $H_{\rm eff}$. It is valid in a theory with cutoff $\Lambda/2$, but its effective "Hamiltonian" $H_{\rm eff}$ depends on the unknown eigenvalue E. However, if $\Lambda/2$ is much greater than E of interest, one can neglect E in comparison to QH_0Q provided that QH_IQ is small in comparison to QH_0Q.

In QCD, which is asymptotically free, one indeed has H_0 as the dominant term in the energy denominator in $H_{\rm eff}$ for small eigenvalues E. In practice, this happens for cutoffs Λ so much larger than the smallest eigenvalues E of physical interest that the corresponding eigenvalue problems are too complex for solving them with required precision. Namely, there are still too many degrees of freedom to account for. One needs to reduce cutoffs considerably further. This issue appears in all approaches to the bound state problem in QCD, not only in the front form of the dynamics.

Even if interactions are sufficiently small, one faces an additional difficulty with eliminating Q-states. Namely, for small interactions one can eliminate the eigenvalue E from a proper effective Hamiltonian in P-subspace in favor of eigenvalues of H_0. Consequently, the denominators analogous to the one that appears above in $H_{\rm eff}$ only contain differences of eigenvalues of H_0, one above $\Lambda/2$ and one below, see Ref. [4]. Unfortunately, such differences can become arbitrarily small near the cutoff $\Lambda/2$ and they generate strong interactions in the effective theory due to the coupling between the states just below and just above the cutoff $\Lambda/2$. This is particularly bothersome when the eigenstates of H_0 near the cutoff are highly degenerate and splitting of the bound state problem into parts below and above the cutoff cannot be accomplished through any simple expansion in powers of the coupling constant.

In any case, when one reduces the cutoff Λ to $\Lambda/2$, and then $\Lambda/2$ to $\Lambda/4$ and so on, the strength of interaction in QCD Hamiltonians increases and, especially if the interaction is attractive, QH_IQ can cancel H_0 and E cannot be ignored no matter how small it is in comparison to the reduced cutoff. In particular, this difficulty concerns bound states, where interactions must prevent free relative motion of constituents from dominating the scene and a spatially compact systems have to be formed. So far, it appears not possible to precisely eliminate the eigenvalue E from the effective dynamics obtained by projecting on sufficiently low energy eigenstates of H_0 to facilitate reliable calculations.

Fortunately, one can use instead a change of basis [8]. Namely, it is possible to define a procedure in which the basis states are rotated in such a way that the matrix elements of H_I vanish between basis states that according to H_0 differ in energy by more than a running cutoff, say λ. The running cutoff is called the energy band width. The name comes from the band diagonal form of the Hamiltonian matrix in the new basis ordered in energy using H_0. Different values of the running cutoff λ correspond to using differently rotated basis states. The rotation is designed not to depend at all on the eigenvalues E one wants to compute.

As a result, one obtains in the rotated basis an effective Hamiltonian matrix eigenvalue problem in which the dependence on cutoff Λ may manifest itself only in the explicit dependence of matrix elements of the new $H_{\rm eff}$. See Ref. [8] for explanation of this result. The two features of similarity that (1) the Λ-dependence becomes explicit before one tackles the problem of solving the eigenvalue problem for $H_{\rm eff}$ and (2) the effective Hamiltonian with small energy band width may not depend on the eigenvalues one tries to find, allow one to discover in advance the required counterterms to the diverging cutoff dependence. A complete set of counterterms defines the set of parameters required for defining the theory which has a finite energy band width λ and no cutoff dependence in the band. In the course of discovering the counterterms and corresponding parameters, one keeps changing the initial Hamiltonian. Eventually, the complete Hamiltonian may have cutoff independent eigenvalues including bound states.

In the case of front form Hamiltonian for QCD, a perturbative version of the similarity renormalization group procedure is outlined in Ref. [9]. Further discussion of computational methods stemming from the similarity renormalization group concept is provided in Sec. C.

C. Renormalization group procedure for effective particles

The similarity renormalization group procedure discussed above in Sec. B can be applied to the problem of describing bound states of quarks and gluons using QCD according to the general computational scheme outlined in Ref. [9] and illustrated in a numerically soluble model in Ref. [10]. Since these works were completed, the method has been applied to various physical systems using the weak-coupling expansion. More recently, similarity has evolved into a computational tool called the renormalization group procedure for effective particles, or RGPEP. In principle, the RGPEP is now defined without a need to refer to some perturbative expansion. Most recent explanation of the RGPEP in terms of an elementary and exactly solvable model for relativistic fermions that interact through a mass mixing term of arbitrary strength in their Hamiltonian, can be found in Ref. [12].

The effective particles can be seen as resulting from a dynamical transformation akin to the Melosh transformation from current to constituent quarks [11]. Namely, the RGPEP transformation changes the bare quanta in a canonical theory to the effective quanta in an equivalent effective theory with a Hamiltonian that has the energy band width λ, see Sec. B and references therein for an explanation of the band. The transformations that change λ form a group.

The effective particles are introduced through a transformation

$$\psi_s = U_s \, \psi_0 \, U_s^\dagger \, , \tag{9}$$

where ψ_s is a quantum field operator built from creation and annihilation operators for effective particles of size $s \sim 1/\lambda$ and ψ_0 is the original quantum field operator built from creation and annihilation operators for point-like bare quanta of a canonical theory. In great brevity, a canonical Hamiltonian density is built from fields ψ_0 and the effective Hamiltonian at scale s is built from fields ψ_s, but without actually changing the Hamiltonian. Thus,

$$H_s(\psi_s) = H_0(\psi_0)\,, \tag{10}$$

which means that the same dynamics is expressed in terms of different operators for different values of s. The coefficients c_s in the expansion of a Hamiltonian in powers of the field operators ψ_s depend on s and the field operators depend on s, but the Hamiltonian is not changing with s. The RGPEP provides an equation for the coefficients c_s as functions of s.

In principle, if one had solved the RGPEP equation for the front form Hamiltonian of QCD exactly, the eigenvalue problem could be written using effective quarks and gluons corresponding to any s. In particular, for s very small, the eigenvalue problem would involve very large numbers of virtual constituents capable of interacting with large momentum transfers up to about the band width $\lambda \sim 1/s$. In contrast, the same eigenvalue problem written in terms of quanta corresponding to a large s, comparable with the size of hadrons, is hoped to take the form of a simple equation that resembles the constituent quark models. To demonstrate mathematically that this is precisely what happens in the RGPEP in QCD is a serious challenge.

---

[1] ILCAC White Paper, arXiv:1309.6333.
S. J. Brodsky, H.-C. Pauli, and S. S. Pinsky, Phys. Rep. **301**, 299 (1998).
M. Burkardt, Adv. Nucl. Phys. **23**, 1 (2002);

[2] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).

[3] E. C. G. Stueckelberg, A. Petermann, Helv. Phys. Acta **26**, 499 (1953).
M. Gell-Mann, F. E. Low, Phys. Rev. **95**, 1300 (1954).

[4] K. G. Wilson, Phys. Rev. **140**, B445 (1965).
K. G. Wilson, Phys. Rev. D **2**, 1438 (1970).

[5] G. t'Hooft, Nucl. Phys. B **61**, 455 (1973).
K. G. Wilson, J. B. Kogut, Phys. Rept. **12**, 75 (1974).

[6] P. A. M. Dirac, Phys. Rev. **139**, B 684 (1965).

[7] R. J. Perry, K. G. Wilson, Nucl. Phys. B **403**, 587 (1993).

[8] S. D. Głazek, K. G. Wilson, Phys. Rev. D **48**, 5863 (1993).

[9] K. G. Wilson *et al.*, Phys. Rev. D **49**, 6720 (1994).

[10] S. D. Głazek, K. G. Wilson Phys. Rev. D **57**, 3558 (1998).

[11] H. J. Melosh Phys. Rev. D **9**, 1095 (1974).

[12] S. D. Głazek, Phys. Rev. D **87**, 125032 (2013).