REFLECTIONS ON THE ROLE OF ASYMPTOTIC ANALYSIS IN PHYSICS

S. A. Fulling

Mathematics Department, Texas A&M University,
College Station, Texas, 77843-3368, U.S.A

1 Introduction

"Asymptotic analysis" is the science and art of systematically revealing the behavior of a mathematical object as some parameter becomes small — usually by expanding in a power series or some suitable generalization of one. It means roughly the same thing as "perturbation theory", but in a generic sense (the small parameter is not necessarily a coupling constant).

Asymptotics has two roles in our thinking. The most obvious is *approximation*: It enables us to find approximate solutions to problems that are too hard to solve exactly. As someone has said, "Every application of mathematics to physics involves a small parameter (though sometimes the small parameter is the degree of contact with experiment)."

But asymptotic thinking is at least as important when it acts in the opposite direction, as *interpretation*. When I begin teaching such material, I ask a class to sketch the graph of the function

$$y = \sqrt{1 + x^2 - \frac{\sin x}{x}}$$

in 30 seconds, without a calculator. Few students will take up the challenge. Then I show them that from what they have already been taught about Taylor series it is easy to see that

$$y \sim x + \frac{1}{2x} \qquad \text{as } x \to +\infty$$

and

$$y \sim \sqrt{\frac{7}{6}}\,|x| \qquad \text{as } x \to 0,$$

and in between y is just going to wiggle a bit, so that one now has a solid picture of how this function behaves.

That is a very precise and mundane example. At the opposite extreme is the grandiose and qualitative issue of "emergent reality", brought out in two famous essays by Philip Anderson [1] and Steven Weinberg [2]. They were debating whether elementary particle physics is really more "fundamental" than other parts of physics. Their eventual agreement is that as the size of a system increases, it becomes *qualitatively different*, and new concepts are needed to describe it. Perhaps someday someone will write a thesis entitled "Derivation of the Navier–Stokes Equations from M-Theory" (including the numerical value of the viscosity coefficient of water); but nobody would ever discover hydrodynamics from just a knowledge of M-theory. So I would compromise the positions of Weinberg and Anderson by saying that, yes, quantum gravity (or something) is fundamental, and although the rest of physics may not be just "device engineering" (as Anderson accused some of his critics of believing), it *is* asymptotics! But it is interpretive asymptotics, like that of $y = \sqrt{1 + x^2 - \frac{\sin x}{x}}$, where the approximation is more interesting and meaningful than the details of the thing being approximated.

An intermediate level of asymptotics is exemplified by *semiclassical approximation*: obtaining classical mechanics, plus small corrections, from quantum mechanics in the limit where $\hbar$ is small compared to the actions S characteristic of the system studied. I would like to concentrate today on a similar and related topic that arose in my early scientific work.

2 Does mathematical form dictate good physics?

Physicists cannot make up their minds on this question. Often we insist that physics is ultimately an experimental science. On the other hand, we feel that the simplest and most elegant theory is usually right. I don't presume to answer the question, but I shall address it from personal experience.

The physical problem involved is field quantization in an expanding universe, as it was investigated by Leonard Parker and me in the early '70s. Our approach had a lot in common with some earlier work by Zel'dovich and Starobinsky [3], but our point of view was different.

We considered a space-time metric

$$ds^2 = dt^2 - a(t)^2(dx^2 + dy^2 + dz^2)$$

and solved the corresponding Klein–Gordon equation as

$$\phi(t, \mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int d^3k \, \left[A_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \psi_{\mathbf{k}}(t) + A^\dagger_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} \psi_{\mathbf{k}}(t)^* \right].$$

The mode functions $\psi_{\mathbf{k}}$ must satisfy

$$\frac{d^2\psi_{\mathbf{k}}}{d\tau^2} = -a^6{\omega_{\mathbf{k}}}^2\psi_{\mathbf{k}}\,, \tag{1}$$

where

$$\tau = \int^t a(t')^{-3}\,dt' \quad \text{and} \quad \omega_{\mathbf{k}}(t)^2 = \frac{k^2}{a^2} + m^2.$$

This much was quite straightforward, and exactly like the theory of the free field in flat space. The problem is that in a time-dependent background all solutions of (1) appear equally worthy. There is no natural definition of "positive frequency" by which to distinguish the physical annihilation operator $A_{\mathbf{k}}$ from, say, $A'_{\mathbf{k}} \equiv \alpha_{\mathbf{k}}A_{\mathbf{k}} + \beta_{\mathbf{k}}A^\dagger_{-\mathbf{k}}$, therefore no clear definition of a vacuum state and of particles. Furthermore, every early attempt to construct such a definition (diagonalizing the instantaneous Hamiltonian, or some equivalent prescription) predicted infinite rates of particle production at most times.

There is one way of invoking mathematics to deal with this problem that didn't lead to much. The reaction of some people at the time was along these lines: "Your analysis is not sufficiently covariant or abstract. You should stop looking at coordinates; you need to find the right Killing vector, the right Lie derivative, the right polarization, the right complex structure, ..., the right second-quantization functor from single particle theories to field theories." In my opinion these viewpoints did not start from the right physics. Often they ignored the clear fact that the field theory describes *particle creation*, at least in certain models where the physical interpretation is unambiguous during certain epochs of time. (If the scale factor $a(t)$ equals a constant for $t < -T$, then $\psi = e^{-i\omega t}$ there and the particle interpretation is clear; but if also $a(t)$ is constant for $t > +T$, then in general $\psi = \alpha e^{-i\omega t} + \beta e^{i\omega t}$ in that epoch, and particles have been created.)

What proved to be the right way to deduce the physical interpretation from the mathematics involved some less abstract, applied mathematics. The essential step was to identify the *effective frequency* of oscillation of each mode at each time. There is an elementary analogy with a *damped* harmonic oscillator: The solutions of the equation

$$\frac{d^2\psi}{dt^2} + 2\gamma\,\frac{d\psi}{dt} + \omega^2\psi = 0 \qquad (0 < \gamma < \omega)$$

are

$$\psi(t) = e^{-\gamma t}\exp\bigl[\pm i\sqrt{\omega^2 - \gamma^2}\,t\bigr].$$

They exhibit damping (as expected) on the time scale γ^{-1} — but also a *frequency shift* of order γ^2 from the unperturbed frequency ω. It is well known in applied mathematics that if one does naive perturbation theory about the old solutions with the old, wrong frequency, one gets a bad answer: one where the error in the approximation grows rapidly with t. (This is called *secular* behavior.)

Much the same thing happens to an oscillator with *time-dependent* frequency. This theory was developed by Fröman [4] and Chakraborty [5] and applied by Parker and me [6, 7, 8] to the cosmological problem. An optimal approximate solution to $\frac{d^2\psi}{d\tau^2} + a(t)^6\omega_{\mathbf{k}}(t)^2\psi = 0$ is

$$\psi_{\mathbf{k}}(t) = \frac{1}{\sqrt{2a^3 W_{\mathbf{k}}(t)}} \exp\left[\pm i \int^t W_{\mathbf{k}}(t')\, dt'\right],$$

with

$$W_{\mathbf{k}}(t)^2 = \omega_{\mathbf{k}}(t)^2[1 + \delta_2(\mathbf{k}, t)\omega_{\mathbf{k}}{}^{-2} + \delta_4(\mathbf{k}, t)\omega_{\mathbf{k}}{}^{-4} + \cdots],$$

where the $\delta_j(\mathbf{k}, t)$ are certain combinations of the time derivatives of $a(t)^6\omega_{\mathbf{k}}(t)^2$ whose precise form is not important for the present discussion. These two functions (with the two choices of $\pm$) are what we used to replace $e^{\pm i\omega_{\mathbf{k}}t}$ in defining annihilation and creation operators. This refinement eliminates the physically implausible prediction that an infinite density of particles is created at intermediate times, which then largely disappears if the expansion of the universe stops.

This *higher-order WKB approximation* is good applied math, but is it a leap of faith to assume that it leads to the correct physics? Let us note several points:

1. The series for W does not usually converge. It must be truncated at some finite point, yielding a nonunique approximation.

2. Thereby the "particles" are defined up to some finite order in powers of $|\mathbf{k}|^{-1}$ and m^{-1}. But as I mentioned previously, during static epochs particles are *precisely* defined, and it turns out that the particle production is *exponentially* small (for a smooth $a(t)$).

3. Therefore, our construction does not define "vacuum" and "particles" exactly. Rather it determines the "right Fock space" containing the physically regular states — those in which field observables such as energy density are *Hadamard renormalizable*, as shown by the work of DeWitt, Christensen, Wald, Kay, Radzikowski, and others.

I conclude that the "adiabatic" definition of particles is not a dogma arbitrarily built on the accidental algebraic form of a particular approximation, but rather a recognition, guided by mathematics, of the "emergent reality" of particle-like behavior in what is fundamentally a field theory (not a particle theory), in a certain limit, that of slow variation of the background geometry.

Recently I have learned of a strange way to invoke very similar mathematics, again to define a physical interpretation. This one, however, strikes me as abuse of the mathematics of asymptotic analysis. To explain this I must first briefly review some more points of background.

First, note that in our problem above, the general solution is a *sum*,

$$\psi = \alpha \frac{1}{\sqrt{a^3 W}} e^{-i\int W\,dt} + \beta \frac{1}{\sqrt{a^3 W}} e^{i\int W\,dt},$$

where the distinguished basis solutions are those that have the right effective frequency. Now it turns out [9] that *any* solution of the differential equation can be written as a single term,

$$\frac{c}{\sqrt{a^3 \mathcal{W}}} e^{i\int \mathcal{W}\,dt},$$

for some function $\mathcal{W}$. However, in general: (1) $\mathcal{W}$ oscillates as rapidly as the solution itself (on the time scale of $\sin(\omega t)$), as opposed to the slow variation of W on the scale of $a(t)$. (2) $\mathcal{W}$ does not reduce to ω whenever $a(t)$ approaches a constant. Although this construction may be useful for some purposes, it clearly has little to do with our strategy of isolating the true frequency of the oscillations.

Second, the WKB approximation has analogues in higher dimensions, applied in particular to wave functions in quantum mechanics (by Maslov [10] and others). In that theory, one seeks a solution of the Schrödinger equation of the form

$$\psi_j = A_j e^{iS_j/\hbar},$$

and one interprets S_j as the action of the classical particle in the sense of Hamilton–Jacobi theory. That is,

$$\nabla S_j = \mathbf{p}_j$$

is the momentum of a beam labeled by j (and then transport along the classical trajectories is used to construct the quantum wave function). In general, ψ is a sum of several such beams (typically four in dimension 2 — see, for instance, Fig. 1 in [11]).

Now to the point: The *Bohm interpretation of quantum mechanics* [12], in its mathematical essence (at least for single-component wave functions) is the demand that one should *always* write the *whole* wave function as $\psi = Ae^{iS/\hbar}$ (just *one* term!) and interpret $\mathbf{p} = \nabla S$ as the momentum of one real physical particle, which randomly finds itself on one of the resulting classical orbits. As for the function $\mathcal{W}$ discussed above, these results are complicated and far from the simple physics of a properly constructed single beam.

The eigenstates $\sin(kx)$ of a one-dimensional particle in a box are *real*, thus p equals 0 in that case, and there is no motion! — even though the particle still has the appropriate energy, $k^2/2m$. The particle just sits in one of the lobes of the function $|\psi|^2 = \sin^2(kx)$.

In higher dimensions there is a famous picture, reproduced by both friends ([12], Fig. 2) and foes [13] of the Bohm theory, which shows how wiggles in the complicated Bohm trajectories create density fluctuations that yield the correct interference pattern in the classic two-slit experiment. These trajectories never cross each other. This means that opening the second slit causes the particle coming from the first slit to steer away from its natural trajectory; it bounces off one of the other trajectories that it might have taken, but didn't.

As you can tell, I am not favorably disposed toward the Bohm theory. It seems to fly in the face of our previous lesson: A good semiclassical analysis needs to start from the identification of the *correct effective frequencies* (generalized to vectors ∇S_j in this case), and a general solution is the superposition of *several* such elementary beams, not just one.

3 How I would have discovered path integrals if Feynman had not already done it 50 years earlier

Now for something (relatively) new. It grew out of my work with Tom Osborn and Frank Molzahn [14, 15, 16, 17] and Davin Potts [18] on asymptotic approximations to fundamental solutions of time-dependent Schrödinger equations.

The *Schrödinger propagator* $K(t, \mathbf{x}, \mathbf{y})$ corresponding to a Hamiltonian operator H is the integral kernel of the operator $e^{-itH/\hbar}$. In other words, the solution of the Schrödinger equation in terms of its initial data is

$$\psi(t, \mathbf{x}) = \int_{\mathbf{R}^d} K(t, \mathbf{x}, \mathbf{y}) \psi(0, \mathbf{y})\, d\mathbf{y}.$$

When there is no potential, K has the well known form

$$K_0(t, \mathbf{x}, \mathbf{y}) = \left(\frac{m}{2\pi i\hbar t}\right)^{\frac{d}{2}} e^{im|\mathbf{x}-\mathbf{y}|^2/2\hbar t}.$$

When

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x}),$$

one can construct an adiabatic, or Wigner–Kirkwood, expansion

$$K \approx K_{WK} = K_0 e^{-\frac{i}{\hbar}\int V(\tilde{\mathbf{x}}(t))\,dt} \exp\Bigg(\sum_{k=1}^{\infty} m^{-k} J_k\Bigg), \tag{2}$$

where the construction of J_k (including the $\int V$ prefactor shown) requires integrations along the *straight line* from $(0,\mathbf{y})$ to $(t,\mathbf{x})$. In contrast, there is a WKB approach to this problem, developed by Van Vleck, Birkhoff, Keller, Maslov, Marcus, Delos, ..., which requires integrations along the *exact classical orbits* from $(0,\mathbf{y})$ to $(t,\mathbf{x})$. The WKB series is an expansion in $\hbar$, while the WK series (2) is clearly an expansion in $\frac{1}{m}$.

Naturally, one tries out these methods on exactly solvable test cases. The simplest case is a constant electric field, $V = \mathbf{E}\cdot\mathbf{x}$. In this case the series (2) truncates after two terms and is exactly correct! Namely,

$$K = K_{WK} = K_{WKB} = K_0 e^{-it\mathbf{E}\cdot(\mathbf{x}+\mathbf{y})/2t} e^{-it^3E^2/24\hbar m}.$$

But this triumph is short-lived. The next simplest case is the harmonic oscillator, $V = \frac{1}{2}m\omega^2x^2$. One finds that the WK series breaks down for $t \lesssim \frac{3}{\omega}$ — that is, in less than half a period! The reason is clear in hindsight: All the exact orbits of this system issuing from point $(0,\mathbf{y})$ refocus at time $t = \frac{\pi}{\omega}$, so a technique that approximates them as straight lines is going to be rather bad. In contrast, the WKB approximation is *exact* for this (purely quadratic) Hamiltonian.

This is bad news. One would dearly love to use straight lines instead of computing exact classical trajectories. In hopes of rescuing the WK method, I contemplated a hybrid asymptotic and numerical approach: Let us evolve the wave function in small time steps, for which the WK approximation is accurate:

$$K(\Delta t,\mathbf{x},\mathbf{y}) \approx K_{WK} = K_0 \exp\Bigg(\sum_{k=0}^{\infty} m^{-k} J_k\Bigg). \tag{3}$$

Iteration to attain the solution at a macroscopic t now requires a spatial integration on each time slice:

$$\psi(t,\mathbf{x}) = \int_{\mathbf{R}^d} d\mathbf{y}_N\, K_{WK}\left(\frac{t}{N},\mathbf{x},\mathbf{y}_N\right)\cdots\int_{\mathbf{R}^d} d\mathbf{y}_1\, K_{WK}\left(\frac{t}{N},\mathbf{y}_2,\mathbf{y}_1\right)\psi(0,\mathbf{y}_1).$$

Then I sat back to look at what I had done, and I realized that I had reinvented the Feynman path integral! Actually, the single-step propagator (3) is

more sophisticated than those in the standard treatments of path integration; the exponent is $J_0 = -\frac{i}{\hbar}\int V(\tilde{\mathbf{x}}(t))\,dt$, plus higher-order terms, as opposed to $-\frac{i}{\hbar}\,\frac{t}{N}\,V(\mathbf{y}_N)$. (See, however, Fujiwara [19] and related papers, e.g., [20].)

Feynman's intention, of course, was to take the limit $N \to \infty$, to *define* quantum dynamics. My strategy was, and is, to take N as *small* as possible consistent with accuracy, so as to get a practical computational scheme. This distinction is the same as that between Riemann's definition of the definite integral and the numerical evaluation of integrals by the trapezoidal rule or Simpson's rule. Note, however, that the path integral is not really needed to *define* the Schrödinger equation. But it is amusing that the path integral can be discovered through the back door as the limit of increasingly accurate WK approximations for a Schrödinger equation assumed from the start. It is as if one defined integration as simply the inverse of differentiation, somehow invented the trapezoidal rule as a way of computing antiderivatives numerically, and then arrived at the Riemann integral by taking the mesh size smaller and smaller in search of better numerical results.

I hope that in the reasonably near future I will be able to present a practical implementation of this scheme and comparisons of its accuracy and efficiency in comparison with the WKB approximation, on the one hand, and conventional partial-differential-equation numerical solution methods, on the other.

Acknowledgments. In addition to the collaborators mentioned in the text, I want to thank Marlan Scully and his associates for acquainting me with the Bohm theory, and J. Avron and Y. Last for correspondence concerning the provenance of the aphorism about small parameters. (They traced it back to J. Zak.)

[1] P. W. Anderson, *Science* **177** (1972) 393.

[2] S. Weinberg, *Nature* **330** (1987) 433.

[3] Ya. B. Zel'dovich and A. A. Starobinsky, *Zh. Eksp. Teor. Fiz.* **61** (1971) 2161 [*Sov. Phys. – JETP* **34** (1972) 1159].

[4] N. Fröman, *Ark. Fys.* **32** (1966) 541.

[5] B. Chakraborty, *J. Math. Phys.* **14** (1973) 188.

[6] L. Parker, *Phys. Rev.* **183** (1969) 1057.

[7] L. Parker and S. A. Fulling, *Phys. Rev. D* **9** (1974) 341.

[8] S. A. Fulling, *Gen. Rel. Grav.* **10** (1979) 807.

[9] P. C. Waterman, *Am. J. Phys.* **41** (1973) 373.

[10] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics*, Reidel, Dordrecht, 1981.

[11] S. K. Knudson, J. B. Delos, and D. W. Noid, *J. Chem. Phys.* **84** (1986) 6886.

[12] D. Bohm, B. J. Hiley, and P. N. Kaloyerou, *Phys. Reps.* **144** (1987) 321.

[13] B.-G. Englert, M. O. Scully, G. Süssmann, and H. Walther, *Z. Naturforsch. A* **47** (1992) 1175.

[14] Y. Fujiwara, T. A. Osborn, and S. F. J. Wilk, *Phys. Rev. A* **25** (1982) 14.

[15] T. A. Osborn and F. H. Molzahn, *Phys. Rev. A* **34** (1986) 1696.

[16] L. Papiez, T. A. Osborn, and F. H. Molzahn, *J. Math. Phys.* **29** (1988) 642.

[17] F. H. Molzahn, T. A. Osborn, and S. A. Fulling, *Ann. Phys.* (N.Y.) **214** (1992) 102.

[18] D. M. Potts, Undergraduate Research Fellow Thesis, Texas A&M University, 1995.

[19] D. Fujiwara, *J. Analyse Math.* **35** (1979) 41.

[20] A. Intissar, *Comm. Partial Diff. Eqs.* **7** (1982) 1403.