Approximations in Physics: A Pedagogic Perspective*

K K Mashood, Arvind Kumar and Anwesh Mazumdar

This article reviews and illustrates some simple aspects of approximations in physics, which should be useful to students at the undergraduate level.

K K Mashood is a faculty member at the Homi Bhabha Centre for Science Education (TIFR). His interests are in physics education research and cognitive science.

Arvind Kumar was a faculty member at the Homi Bhabha Centre for Science Education (TIFR) till his retirement in 2008. His main interests are theoretical physics and physics education.

Anwesh Mazumdar is a faculty member at the Homi Bhabha Centre for Science Education (TIFR). His research interests include stellar astrophysics and physics education.

1. Introduction

Approximations are encountered all too frequently in physics. Why are they so necessary? The simple reason is that real systems in nature are far too complex for physics to analyze them exactly in every detail. A macroscopic system (i.e. a system of ordinary length, mass and time scales), in general, will need a large number of variables to describe it fully, even if we ignore for the moment its microscopic constitution. For example, a rocket that is moving against gravity by jet action due to the exhaust of gases produced by some combustible substances inside has mixed mechanical, chemical and thermodynamic description, and it would be difficult if not impossible to analyze the system in its full detail. Instead, we handle the problem by ignoring some features and focusing on others, depending on our interest.

For example, if our main interest is the motion of the rocket as a whole, we need to know only some mechanical variables (the original mass, the rate of mass depletion, the exhaust velocity relative to the rocket and gravitational acceleration.). The problem is then reduced to the standard variable mass problem in mechanics. Further approximations may be made to simplify the treatment such as taking exhaust velocity and gravitational acceleration to be constants; and so on. On the other hand, if our interest is in, say, the explicit calculation of the mass depletion rate and exhaust velocity, we will focus on the thermodynamics and chemical ki-

*Vol.25, No.7, DOI: https://doi.org/10.1007/s12045-020-1009-8

Keywords
Approximation, derivations, modelling, pedagogy.

netics of the fuel and ignore for the moment the overall motion of the system.

This strategy of truncation of reality and approximate description of systems in nature is generally known as 'modeling', and it has worked wonderfully in physics. A model is not the real system; it aims to mirror approximately features or aspects of the system that interest us. Its purpose is usually to render a problem analytically or numerically tractable. But sometimes modeling goes beyond this aim. A simple model of a real system can occasionally provide insights that a more complete treatment might not. A striking example in history is Einstein's 1905 calculation of (volume dependence of) entropy of thermal radiation inside a closed enclosure. To focus on its non-classical features, Einstein chose to handle the problem in the short wavelength approximation (where the classical Rayleigh-Jeans law was known to fail and the simple Wien's radiation law worked), and arrived at the corpuscular (photon) picture of radiation, which he used to explain the puzzling features of the photoelectric effect. Had Einstein not made the short wavelength approximation but instead worked with the exact and more involved law of black body radiation (derived earlier by Planck in 1900), it is likely he would have got embroiled with the messy details of its entropy calculation and missed the fundamental insight he had obtained from an approximate simple treatment.

This strategy of truncation of reality and approximate description of systems in nature is generally known as 'modelling', and it has worked wonderfully in physics. A model is not the real system; it aims to mirror approximately features or aspects of the system that interest us.

Closely related to approximations are the heuristic strategies of idealization and thought experiments. An ideal gas does not exist for it cannot undergo phase transition that all real gases do. Yet it is an extremely useful construct to describe a real gas; the ideal gas law is found to be an excellent approximation for any dilute (i.e. low density) gas at high temperatures far away from its critical temperature. Thought experiments are nothing but exploration of systems and processes under ideal (if impractical, but, in principle, possible) conditions. Their enormous role in arriving at radical new concepts and laws is evident in history in the thought experiments of, for example, Galileo (involving frictionless inclined planes) and Einstein (involving accelerated elevators), which led respectively to the law of inertia and the Principle of Equivalence.

The relation/distinction between approximation, idealization, model construction, thought experiment, etc., has been subject to considerable philosophical scrutiny. Since this article is pedagogic in nature, we shall be content to only give some references to the philosophical literature [1, 2, 3] and use the term 'approximation' below in a broad sense that subsumes all other related terms above.

In this article we attempt to delineate the different elementary aspects of approximations in physics that are part of common knowledge of teachers and experts.

Any physics text or instructional course at the college/university level is replete with derivations and results involving approximations. Experience shows that most students even at fairly senior stages do not appreciate the rationale of this practice, but rather view it as a bag of adhoc tricks cleverly adapted to different contexts to get to the 'expected' results. In this article we attempt to delineate the different elementary aspects of approximations in physics that are part of common knowledge of teachers and experts. We believe that clearly identifying and articulating them explicitly should be pedagogically useful. It may be noted that we are dealing here with approximations in theoretical derivations/results in physics; the issues of numerical approximations and measurement-related errors are not addressed.

2. Aspects of Approximations

Reviewing different theoretical derivations and results in physics, we identify some salient and recurring aspects underlying the approximations employed in them. We illustrate them by a few familiar examples. We believe these aspects cover most of the approximations encountered in university level physics.

2.1 *An approximation in physics usually entails some dimensionless quantity, say α, being considered very small; $\alpha \ll 1$.*

This rather obvious point is not always emphasized in our instruction. The result is that we frequently encounter statements like "If the zero of potential energy is taken at the earth's surface, the potential energy of a body at height is mgh for small h; an anharmonic oscillator is approximately harmonic for small displacement x; the Uncertainty Principle $\Delta x \Delta p \geq \hbar$ has no role at the macro level since $\hbar$ is too small, the usual treatment of Zeeman effect is correct for weak magnetic fields" and other similar assertions.

The underlined phrases above are evidently flawed since a suitable change of units can make the quantities as large as one wishes. The first statement is easily corrected; replace h by the dimensionless h/R, where R is the radius of the earth. The second needs to be phrased more carefully. If a is the force constant in the harmonic term ax, and b the anharmonicity constant occurring in the term say bx^3, the approximation is valid when the amplitude A is much smaller than $\sqrt{a/b}$ or the dimensionless ratio $A/\sqrt{a/b}$ is much smaller than unity. Note that the approximation is properly stated in terms of maximum displacement (A), and not x, which is a variable and hence will always satisfy the approximation criterion for some range of values.

Note that the approximation (of an oscillator being simple harmonic) is properly stated in terms of maximum displacement (A), and not x, which is a variable and hence will always satisfy the approximation criterion for some range of values.

The third assertion can be debated in physics, but will at least make sense dimensionally if we say $\hbar$ is much smaller than the typical values of macroscopic angular momenta. The fourth phrase is intended to mean that the magnetic interaction energy of the electron is small compared to its bound state energy in the electric field of the nucleus. In some contexts the same dimensionless ratio α may be $\ll 1$ in one limit and $\gg 1$ in another limit. For a radiating charge-current source, the near field approximation is characterized by $\alpha = r/\lambda \ll 1$, and the far field approximation by $\alpha = r/\lambda \gg 1$, where r is the distance of the observation point from the source and λ the wavelength of the radiation (the latter case may, of course, be described by $\beta = 1/\alpha \ll 1$) . In both cases the source dimension d is small compared to r and λ.

It might be argued that there is no need to be so pedantic and it is obviously understood (though not explicitly stated for brevity) that a dimensional quantity is to be compared with the relevant quantity of the same dimension. However, the relevant quantity to compare with, even for simple cases, may not be obvious to many students; for some cases it may in fact require closer examination of the problem. It is therefore best to clearly state the appropriate dimensionless quantity when dealing with an approximation.

2.2 *Approximations may entail more than one dimensionless ratios, say α and β, to be $\ll 1$.*

Often, there are two (or more) independent quantities being considered small. For example, a real gas approaches an ideal gas when both the effects of intermolecular forces and the finite volume of the molecules in the gas equation are neglected. For a Van der Waals gas, this approximation is best described in terms of two dimensionless ratios: $b/\bar{V} \ll 1$ and $\bar{V}T \gg a/R$, where a and b are the gas-specific constants in the Van der Waals eq. $P = \frac{RT}{\bar{V}-b} - \frac{a}{\bar{V}^2}$; connecting pressure P to molar volume $\bar{V}$ and absolute temperature T.

Another familiar example is the neglect of Coriolis and centrifugal accelerations compared to the acceleration due to gravitational force (g) on the surface of the earth spinning with angular speed ω. In the simpler case of equatorial motion with speed u, we can write $a = g \pm 2\omega u - \omega^2 R$ (the sign of the second term depending on the direction of motion), whence it follows that the validity of the approximation can be phrased in terms of two dimensionless ratios: $\alpha = \frac{\omega^2 R}{g} \ll 1$ and $\beta = \frac{\omega u}{g} \ll 1$. These are general conditions for neglecting inertial forces compared to the gravitational force on a spinning planet. For the case of earth, the first is true for the known values of the parameters, while the second limits the validity of the approximation in terms of the speed of the body.

Examples abound and it should be useful to be alert about this in different derivations in physics. For instance, the Newtonian limit of Einstein's theory of gravity involves two approximations, the non-relativistic approximation $v/c \ll 1$ and the weak field approximation $h_{ik} \ll 1$ (here the gravitational metric tensor is given by $g_{ik} = \eta_{ik} + h_{ik}$, where h_{ik} denotes the small departure

It is not enough to say that a certain dimensionless ratio α is taken to be small. To quantify the statement, we need to state to what power of α is our calculation correct, or equivalently, which power of α onwards are we ignoring in the calculation.

from the Minkowskian metric tensor η_{ik}) [4]. If we use only the second and not the first, the result will be the linearized theory of gravitational waves. If we use only the first and not the second, it would not be much useful since a strong gravity regime should make all motion relativistic, whatever the initial conditions.

Occasionally, a derivation may involve two dimensionless ratios that involve one common quantity. A well-known derivation in optics gives the separation between two successive maxima in a double slit experiment equal to $D\lambda/d$ (d being the separation between the slits, λ the wavelength of light and D the distance of the screen from the slits). The derivation involves the approximations $\lambda/d \ll 1$ and $d/D \ll 1$. Thus the same quantity d is comparatively large in one approximation and small in the other. There is, of course, nothing inconsistent, but it is worth noting, for this sort of thing comes up frequently.

Merely identifying two independent dimensionless ratios say α and β in a problem does not complete the task. We need to see what function(s) of these is (are) relevant for the approximations we have in mind. Returning to the earth's example, we could define another dimensionless ratio, say $\gamma = \beta/\alpha = u/\omega R$. Saying now both α and γ are much less than unity in a certain situation implies that not only are the two inertial forces negligible but also that the Coriolis force is much less than the centrifugal force. Conceivably, we might then ignore the Coriolis force altogether but retain the centrifugal force to the lowest order.

2.3 *The notion of 'order of approximation' is important for proper quantitative characterization of an approximation.*

It is not enough to say that a certain dimensionless ratio α is taken to be small. To quantify the statement, we need to state to what power of α is our calculation correct, or equivalently, which power of α onwards are we ignoring in the calculation. If the approximation involves neglecting terms of order α^{n+1} and higher powers, the approximation is correct up to order α^n. To be able to say this, we need to expand every function involving α in a power series using the binomial series or other familiar series expansions of trigonometric, hyperbolic, exponential and logarithmic functions. [Remark: Incidentally, this also means that the argument of all polynomial and transcendental functions appearing anywhere in a physical equation must always be dimensionless. Only a monomial can have an argument that carries some dimension.]

For example, the correct relativistic formula for kinetic energy of a particle of rest mass m_0 and speed v,

$$K = m_0 c^2 \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right)$$

reduces to the Newtonian relation $\frac{1}{2} m_0 v^2$ if we ignore terms of fourth power in v/c and higher. It is easily seen that the binomial expansion in this case has no cubic power in v/c. Thus the Newtonian relation is a correct approximation up to $(v/c)^3$, not up to $(v/c)^2$, as some might be prone to say. To take an example, suppose the rest mass energy of a particle is 2 MeV. Further let the speed v of the particle be $0.1c$, where c is the speed of light. Then $(v/c)^2 = 0.01$. The Newtonian formula then gives kinetic energy of the particle to be 0.01 MeV. The first term neglected in the formula is $m_0 c^2 (3/8)(v/c)^4$ which in our example is three fourths of 0.0001 MeV. Clearly, the Newtonian answer is correct up to three not two decimal places. It is better to write it as 0.010 MeV. Likewise, putting $\sin\theta = \theta$ in the equation of motion of a simple pendulum (which makes the motion simple harmonic) is correct up to θ^2 not just up to θ, since the terms being ignored are θ^3 and higher. This is why even for a moderately large angle of, say, 15 degrees ($\pi/12 \approx 0.2618$), the approximation is reasonably good ($\sin(\pi/12) \approx 0.2588$), i.e. correct up to two decimal places. [Remark: Saying to what order is an approximation valid is partly a matter of definition. Some may like to equate it to the last power (of the dimensionless ratio) retained in the calculation, while here it is equated to one less than the first power ignored. The latter definition makes more sense as explained in this paragraph.]

The argument of all polynomial and transcendental functions appearing anywhere in a physical equation must always be dimensionless.

If there are two dimensionless ratios α and β, an approximation may be correct up to different orders in α and β. For example, the simplest version of Bohr model for hydrogen atom regards the nucleus to be an infinitely massive point object — its role is only to provide the Coulomb force for the orbiting electron. If we wish to estimate the effect of finite mass and finite size of the nucleus, two dimensionless ratios $\alpha = m/M$ (m being the mass of electron and M the mass of nucleus) and $\beta = R/a$ (R being the radius of nucleus and a the Bohr radius) enter the calculation. Now, an elementary treatment shows that the leading correction due to finite size of the nucleus is of second order in R/a. On the other hand, the effect of finite mass of the nucleus (regarded as a point object) is exactly known: the energy formula is simply modified by a multiplicative factor $(1 + m/M)^{-1}$. Now suppose for the purpose of this discussion, R/a is of the order of $(m/M)^{3/2}$. A numerically consistent approximation including both the effects would then be of second order in R/a and third order in m/M. In general, a consistent approximation may involve different orders of terms in the two ratios α and β.

2.4 *Consistency of approximation up to a certain order requires that all terms contributing up to that order (and no higher) be included.*

This theme follows naturally from the notion of the order of approximation discussed above. A nice example can be found in the problem of the fine structure of the hydrogen atom. As we know, Bohr's (non-relativistic) model yields highly degenerate energy levels of the hydrogen atom. The observed hydrogen spectrum, however, shows fine structure which means that this high degeneracy must be split due to some corrections to the Bohr model. One obvious correction is that due to relativistic effects in the energy of levels where the electron is quite close to the nucleus. An elementary treatment in quantum mechanics determines this relativistic correction to order v^2/c^2 times the energy of the original model. However, this does not quite match the observed fine structure. Something is missing.

Now alkali spectra also show fine structure. Since the valence electron is far away in this case, this cannot be due to a relativistic correction. It is now known that there is another effect: the spin-orbit coupling, which leads to corrections of order v^2/c^2. This correction arises because the intrinsic magnetic moment associated with electron spin interacts with the magnetic field seen by the electron in its rest frame due to the orbiting motion of the nucleus (proton in the case of hydrogen).

The two effects when added up for the hydrogen atom still do not agree with the observations. It turns out there is yet another relativistic correction of the same order: that due to the so-called Thomas precession. This arises from the special relativistic kinematic effect that a fixed spin vector in the electron's rest frame precesses in the proton rest frame: the laboratory frame. When all these corrections of the same order are added up, there is good agreement with experiment. [Remark: The history of the idea of spin-orbit coupling is fascinating. After Bohr's model, Sommerfeld introduced relativistic effects in the frame-work of *old* quantum theory and produced a fine structure formula that worked well not only for hydrogen spectrum but also where it should not have worked (!): the alkali spectra. This led to search for a physical effect in alkali atoms that will produce corrections of the same order as relativistic corrections in hydrogen atom. After some twists and turns (that included Pauli's near miss of the discovery of electron spin), it was surmised that the interaction energy due to the spin magnetic moment of electron was the source of this effect, but initially there was confusion since the magnetic field was nowhere to be seen. It was soon realized that transforming to the the electron's rest frame, there *is* a magnetic field with which its spin magnetic moment interacts. This effect when corrected for by Thomas precession finally gave the correct spin-orbit interaction.]

It is one of the many triumphs of the Dirac equation discovered in 1928 that the Dirac Hamiltonian for hydrogen atom systematically reproduces to the order v^2/c^2 these different relativistic corrections, which were obtained in a patchy way earlier [5]:

$$\left[\frac{p^2}{2m} + V - \frac{p^4}{8m^3c^2} + \frac{1}{2m^2c^2}\frac{1}{r}\frac{dV}{dr}\vec{S}.\vec{L} - \frac{\hbar^2}{4m^2c^2}\frac{dV}{dr}\frac{\partial}{\partial r}\right]\psi(\vec{r}) = E\psi(\vec{r}) \quad (1)$$

To get corrections in a theory up to a given order, we must make sure to include all possible effects that contribute to that order. It also does not make sense to calculate one effect to higher orders than the rest.

The first two terms represent the usual non-relativistic Hamiltonian, the next term is the kinematic relativistic correction obtained by expanding the kinetic energy relation to include the correction of order v^2/c^2 times the the main non-relativistic term $p^2/2m$.

The fourth term involving the intrinsic spin $\vec{S}$ and orbital angular momentum $\vec{L}$ operators, besides the radial derivative of the spherically symmetric potential energy V, is found to be precisely the Thomas precession corrected spin-orbit coupling term obtained before by heuristic methods; the last term has no classical analogue. It is easily checked that the fourth and fifth terms are again of order v^2/c^2 times the main term. To do this checking, take S and L to be of order $\hbar$ which is of order rp. Also note $\hbar\frac{\partial}{\partial r}$ is like p.

Such examples are scattered throughout the subject. They caution us that to get corrections in a theory up to a given order, we must make sure to include all possible effects that contribute to that order. It also does not make sense to calculate one effect to higher orders than the rest. Thus in the above example, we could calculate kinematic relativistic correction to still higher order since the exact formula for relativistic kinetic energy is known; but it is no use if the remaining terms are also not corrected to that order.

The importance of the order of approximation cannot be overemphasized. Competing theories may have the same predictions up to a certain order in which case experiments have to be refined enough to go to the next higher order to decide one in favour of the others. For example, though in principle, the notion of 'ether' could have been assessed by experiments measuring first order in v/c effects, it needed the measurement of possible v^2/c^2 effects (as in the Michelson Morley experiment) to accept or reject it, since a null result up to first order could be accommodated, albeit unsatisfactorily, without abandoning the notion of ether [6]. In

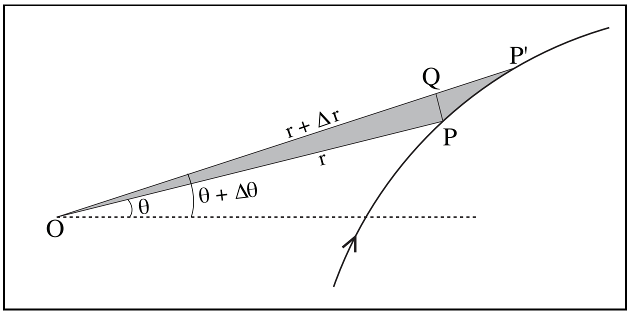

Figure 1. Schematic of a particle moving in a plane.

other words, a theory may be falsifiable only in a certain order of the relevant (dimensionless) parameter, not in a lower order.

A calculation that is algebraically consistent in terms of 'order of approximation' should be usually numerically consistent too. Note, however, that there is an unwritten assumption that the coefficients of terms of successive powers of the dimensionless ratio are of the same order of magnitude numerically. However, this may not always be so. For example, it is possible that, say, two terms contributing in a lower order subtract while those in the next order add up, resulting in the situation that the higher order term may numerically be about the same, or even greater than the lower order term. Such situations need to be handled on a case by case basis.

2.5 *The infinitesimal quantities introduced in a calculus based derivation amount to no approximation, since they go to zero in the limit.*

This point is best illustrated by an example. Consider a particle moving in a plane (see *Figure* 1). The angular momentum of the particle is by definition $\vec{L} = \vec{r} \times \vec{p}$. For planar motion, the direction of $\vec{L}$ is normal to the plane, 'up' or 'down' depending on the choice of origin and the sense of traversing the curve in the plane. In the figure, $\vec{L}$ is normal coming out of the plane. In plane polar coordinates (r, θ, with origin at O) its magnitude is given by $L = mr^2 \frac{d\theta}{dt}$. Let us now calculate the area swept by the particle (shaded in *Figure* 1) per unit time.

The particle moves from the point P to P′ in a time interval Δt. Draw PQ $\perp$ to OP′. For small Δt, P and P′ are quite close and the area covered by the portion PQP′ is negligible compared to the area of the $\triangle$OPQ. Therefore, area swept in time Δt is equal to area of $\triangle OPQ = \frac{1}{2}$(OQ)(PQ).

Now OQ $= r\cos\Delta\theta$ and PQ $= r\sin\Delta\theta$, which implies area of $\triangle OPQ = \frac{1}{4}r^2\sin 2\Delta\theta = \frac{1}{2}r^2\Delta\theta$, since $\Delta\theta$ is small. Therefore, area swept in time Δt equals $\frac{1}{2}r^2\Delta\theta$ and the areal velocity A is,

$$A = \frac{1}{2}r^2\frac{d\theta}{dt} \tag{2}$$

Thus we get,

$$L = 2mA \tag{3}$$

Is this an exact relation? Having learnt the notion of order of approximation a student might suspect that this result is approximate, since we took $\sin 2\Delta\theta = 2\Delta\theta$ and also ignored the area of the portion PQP′. After all, replacing $\sin\theta$ by θ in pendulum motion gives an approximate not an exact result. A moment's reflection shows that here the argument of sine goes to zero in the limit. We retain the first order infinitesimal $d\theta$ since we are finding the area swept in first order infinitesimal dt. The order of infinitesimal must match on both sides. What then justifies the neglect of portion PQP′? Arguing along the same line the area of PQP′ is of the order $rdrd\theta$. This is a second order infinitesimal (a product of two first order infinitesimals), one order higher than dt and the area of OPQ. The ratio of this area to dt goes to zero in the limit, so neglecting it leads to an exact, not an approximate result.

In calculus based derivations in physics, we deal with quantities that would be going to zero in the limit (in the example above, areas of both $\triangle$OPQ and region PQP′ go to zero in the limit); yet we retain some and neglect others. This can be bewildering to a beginning student. The notion of order of infinitesimal is very helpful to understand what is going on in a derivation, and why at the end the result is exact, subject of course to any simplifying assumptions of the model for which the calculation is done.

2.6 *A reasonable estimate of the correction due to several effects together is usually obtained by simply adding leading corrections due to each individual effect calculated by neglecting the other effects.*

This point is again best understood by an example. In Newtonian theory a planet moves in a fixed orbit (ellipse) in a plane under the gravity of the star. However, if the effect of the gravity of other planets and certain other small effects are taken into account, there is a small correction to the simple theory that results in a precession of the orbit. For the planet Mercury in the solar system, this effect is about 531.1 seconds of arc per century. As is well-known, this theoretical prediction had a small discrepancy from the observed precession of 574.1 seconds of arc (after allowing for the precession of the equinoxes).

Now when Newton's theory is modified to Einstein's theory of gravity, how should one proceed to examine if the new theory explains or does not explain the discrepancy? Should one apply the new theory to the complicated problem of Mercury under the gravitational influence of the star and the other planets? Common sense dictates that since the dominant gravitational influence on Mercury is by the Sun, the effect of the new theory should be first calculated for the simpler situation where the presence of other planets and other small effects is ignored. Remarkably, as was first calculated by Einstein, this gives a correction of about 43 seconds of arc which simply added to the earlier correction explains the observed precession within experimental errors.

The common sense remarked above is nothing but a feel about the order of approximation. If one imagines (unrealistically) doing a calculation in the new theory taking into account the other planets as well, the additional correction to the leading correction would go to higher order. Thus to calculate the total leading correction due to the two effects (planetary perturbations in Newtonian theory, and modifications of Newtonian theory itself), we calculate the correction due to each effect ignoring the other effect, and simply add up the corrections.

To calculate the total leading correction due to the two effects (planetary perturbations in Newtonian theory, and modifiations of Newtonian theory itself), we calculate the correction due to each effect ignoring the other effect, and simply add up the corrections.

We do not need to go to this advanced example to appreciate the point. A simple pendulum in the small amplitude approximation has a period that is independent of amplitude, if we ignore air resistance. Now suppose we wish to estimate the correction due to large amplitude as well air resistance. As a first approximation, it is sensible to estimate the leading correction due to non-negligible A/l (A: amplitude, l: length of pendulum) ignoring air resistance, estimate the correction due to air resistance ignoring the first effect; and add up the two corrections.

To examine this point a little better, suppose α and β are the dimensionless perturbation parameters for the two effects, and the leading corrections are in the first order. Then the coupling correction of the two effects is likely to be of order $\alpha\beta$ while the sum of the uncoupled corrections is $\alpha + \beta$. The coupling correction is then of a higher order and can be ignored. We can rule out the possibility of the coupling correction to be of the type $\sqrt{\alpha\beta}$ on the reasonable (though perhaps not rigorously provable) grounds that a perturbed solution is unlikely to contain non-analytic functions like square root terms in the basic perturbation parameters. However, if it so happens that the leading corrections appear in second order only (i.e. are of order α^2 and β^2),it is conceivable that the coupling term may be in the same order ($\alpha\beta$), which then cannot be ignored. Excepting such special circumstances, the procedure stated in the theme seems sensible, especially if the two corrections do not differ greatly in their orders of magnitude. It, however, needs more rigorous justification than is given here, depending on the problem at hand.

3. Concluding Remarks

The preceding discussion was aimed to highlight the fact that making valid approximations in physics is a fairly systematic exercise; there is a method in what beginning students might perceive to be adhoc. Admittedly, our entire discussion is at a rather elementary level which misses out several important dimensions of the subject. One important question is: does every approximation in physics involve a dimensionless quantity being considered small? To deal with this question, let us consider several general contexts of approximations in physics:

Making valid approximations in physics is a fairly systematic exercise; there is a method in what beginning students might perceive to be adhoc.

1. The simplest scenario is when an exact result of a theory is available but we need it only approximately. In this case it seems evident that the exact result will need to be expanded as a series (finite or infinite) in terms of some dimensionless quantity and truncated up to the desired order. Examples are the general formula for relativistic kinetic energy or the exact energy formula for hydrogen atom given by Dirac equation; the concerned dimensionless quantities being v/c and fine structure constant α, respectively.

2. Another situation is when the given problem is not exactly solvable but a closely related one is. This is the scenario for classical and quantum perturbation theories. Examples abound and need not be given here; they can be looked up in standard texts. In this case, the perturbation expansion again involves a dimensionless quantity that relates to the ratio of the perturbation of the quantity to the unperturbed quantity (say, change in the Hamiltonian to the original Hamiltonian).

3. But not all approximations are of these kinds. For example, a frequently encountered approximation is the adiabatic approximation applicable when something (boundary conditions or some part of the system) changes slowly, on a time scale much greater than the time scale of the main dynamics of the system. A harmonic oscillator whose equilibrium point moves slowly over a time much larger than its time period is an example. The well-known Born–Oppenheimer approximation used in molecular physics is another example; here the nuclear motion is much slower than the electronic motion. At the opposite end is the so-called ‘sudden approximation’ wherein the time scale of the change in boundary conditions, etc., is much smaller than that of the system’s dynamical time scale.

This is so for numerous contexts of approximations in physics–we may characterize their validity in terms of some dimensionless quantity but it would not be easy to use it for improvement of the approximations.

The relevant dimensionless quantity that characterizes such approximations is obvious, namely the ratio of the two time scales. But the approximations may not have a clear framework to obtain corrections in terms of this quantity. This is so for numerous contexts of approximations in physics–we may characterize their validity in terms of some dimensionless quantity but it would not be easy to use it for improvement of the approximations.

4. An example of a very different kind of approximation is the calculation of the number of microstates (Γ) of an ideal gas in the macrostate (N, V, $E - d/2$ to $E + d/2$); N is number of molecules, V the volume, and E the mean energy with a small width d. This statistical mechanical problem actually requires the calculation of $\ln\Gamma$. We might initially expect the dimensionless quantity d/E to be the significant determinant of the calculation. It turns out, however, that its value hardly matters for $\ln\Gamma$! The basic reason can be traced to the fact that the number of states between 0 to E increases with E at an exceedingly high rate (since N is enormously large); for details, see [7]. Thus if we insist on a small dimensionless quantity in this example, it is $1/N$. But there is hardly a need to consider corrections in terms of this, since N is of the order of Avogadro's number.

5. Finally, a naive application of the idea of 'approximation in terms of a small dimensionless quantity' can sometimes lead to fundamentally misleading conclusions. For example, we might wrongly think that quantum equations reduce exactly to classical equations in the $\hbar$ going to zero limit or that there can be no macroscopic quantum effects since the dimensionless ratio of $\hbar$ and typical macroscopic angular momenta is exceedingly small.

These remarks caution you that this article does not cover the entire variety of approximations used in physics. But the aspects covered here should be useful in a great majority of situations and help a student to be better prepared to handle the more advanced contexts of approximations.

Another dimension that we have not touched upon here involves issues when approximate results of a theory are compared with experimental results or when experiments are used to choose the best candidate among alternative theories. The estimation of errors due to known and random factors in advanced experiments involves sophisticated techniques of error analysis. Likewise, any theoretical prediction is subject to errors both due to factors internal to the theoretical model (i.e., approximations made in the deductions of the model) and external to it (i.e., the assumptions of neglecting certain physical effects in the very construction of the model.) In theories involving complex and large scale numerical computation, calculation of margins of error is yet another complicated task. All this makes the validity of any claim of theoretical agreement with experiments a highly non-trivial issue. Yet, this has been accomplished to an astonishing degree in several domains of physics.

Awareness of nuanced aspects of approximation can help making it less vexatious to beginning university students, and can also go some way in their appreciating the nature of physics practice. This ties up with the discourse in science education that emphasises making expert's implicit knowledge explicit in instruction.

The validity of approximations has been an issue in the history of physics. Boltzmann, for example, distinguished between two types of approximations: first, in which the neglected terms due to some assumptions in the theory could be of lower order than the final result for appropriate physical systems; and second, in which the neglected terms were of the same order as the terms that led to the final result. The logically untenable second kind of approximations are still made on grounds of simplicity and tractability; see the discussion in [[8]]. Thus in physics practice, exceptionally, approximations may not have proper justification and other reasons, primarily the need to make analytical progress in a theory or to arrive at some insights, are invoked to still pursue with them.

To conclude, though the topic of approximations in physics is deep and all-encompassing, this article elaborates on some of its most elementary aspects, awareness of which can help making it less vexatious to beginning university students, and can also go some way in their appreciating the nature of physics practice. In this sense, it ties up with the current discourse in science education that emphasizes making the experts' implicit knowledge and practices explicit in instruction. For further investigation of students' hurdles in this topic, a diagnostic tool based on the different aspects of approximations in physics discussed here has been developed at our institution and is being administered on a large sample of students across the country. The results of this empirical study will be reported in due course.

Acknowledgements

This work was carried out under the National Initiative on Undergraduate Science (NIUS) programme undertaken by Homi Bhabha Centre for Science Education, Tata Institute of Fundamental Research, Mumbai, India. We acknowledge the support of the Govt. Of India, Department of Atomic Energy, under Project Identification No. RTI4001.

Suggested Reading

[1] **J D Norton, "Approximation and idealization: Why the difference matters",** ***Philosophy of Science*****, Vol.79, No.2, pp.207–232, 2012.**

[2] **D P Portides, "The relation between idealisation and approximation in scientific model construction",** ***Science and Education*****, Vol.16, Nos.7–8, pp.699–724, 2007.**

[3] **C Liu, "Approximation, Idealization, and Laws of Nature",** ***Synthese*****, Vol.118, No.2, pp.229–256, 1999.**

[4] **M P Hobson, G P Efstathiou and A N Lasenby,** *General relativity: an introduction for physicists*, **Cambridge University Press. New York, 2006.**

[5] **L I Schiff,** *Quantum Mechanics*, **p.482, McGraw-Hill, New York, 1968.**

[6] **A I Miller,** *Albert Einstein's Special Theory of Relativity*, **pp.21, 41, Addison-Wesley, Massachusetts, 1981.**

[7] **R K Pathria and P D Beale,** *(Statistical Mechanics*, **pp.11–15, Elsevier, New Delhi, 2011.**

[8] **J L Ramsey, "Towards an expanded epistemology for approximations",** ***In PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*****, Vol.1992, No.1, pp.154–164, Philosophy of Science Association, 1992.**

Address for Correspondence
K K Mashood[1]
Arvind Kumar[2]
Anwesh Mazumdar[3]
Homi Bhabha Centre for Science Education Tata Institute of Fundamental Research
V. N. Purav Marg, Mankhurd
Mumbai, 400088, India.
Email:
[1]mashood@hbcse.tifr.res.in
[2]arvindk@hbcse.tifr.res.in
[3]anwesh@hbcse.tifr.res.in