Back of the Envelope Calculations

One of the purposes of a physics education is to teach students the most formal and intellectually rigorous way to derive conclusions and predictions from the principles of physics. But often such high levels of rigor are not necessary to obtain a qualitative understanding of a system or even a reasonable numerical answer to a physical question. At much lower levels of rigor than what is taught in physics classes, exist back of the envelope calculations. The aim of these notes is to argue for the utility of these calculations and discuss good practices for setting them up.

1 Introduction: The Rings of Saturn

Let's start with a physics question:

> Are the rings of Saturn of uniform mass density?

This question might not immediately strike one as a physics question. First, there are no defined variables or explicit numerical quantities, and, perhaps most egregiously, there is no reference to an equation. Of course, in the wild of the proverbial 'real world' this is how all physics questions come to us: Naked and devoid of any semblance to those precise theories which supposedly describe this world. Part of the process of answering such questions requires transforming the question into one which makes contact with the terminology of these theories. For the question of whether Saturn's rings are uniform, we could restate it as whether it is possible for uniform rings around a planet to exist as long as the rings of Saturn have existed. Namely, do the uniform rings around a planet constitute a stable gravitational system?

In order to answer this new question we could set up an orbital model of a planet and a uniform ring, work through a Newtonian or Lagrangian mechanics analysis, and then derive the conditions of stability. This would be a rather exact analysis of this problem, but we could find the same answer at a much lower level of detail. Given that the gravitational force varies as $\sim 1/r^2$, and the amount of mass in an arc-like region varies as $\sim r$ (where r is the distance from the planet to the mass), the force from such a mass region varies as $\sim 1/r$, and so there will always be an imbalance of forces on a planet displaced from the center of a uniform ring. Therefore, the system is unstable, and Saturn's rings (which form a stable system with the planet) cannot be uniform.

More precisely—but only slightly more so—consider the situation shown in Fig. 1. We can determine the stability of a system by determining whether, upon deviations, degrees of freedom in the system return to their equilibrium configurations. For a planet and uniform ring, that equilibrium configuration consists of the planet at the center. Say we were to move a mass M_P planet away from the center of a ring with mass density λ, so that the planet was a distance r_I away from one side of the ring and a distance $r_{II} > r_I$ away from the other side. Then, for ring arc-lengths subtending the same angle $\Delta\phi$, the mass on one side would be $\Delta m_I = \lambda \Delta\phi\, r_I$ and the mass on the other side would be $\Delta m_I = \lambda\Delta\phi\, r_{II}$. The net-force on the planet along the axis connecting the two relevant sides of the ring could then be approximated as

$$F_{\text{net}} \sim G\frac{M_P \Delta m_I}{r_I^2} - G\frac{M_P\Delta m_{II}}{r_{II}^2} \sim GM_P\lambda\Delta\phi\left(\frac{1}{r_I} - \frac{1}{r_{II}}\right) > 0, \tag{1}$$

where we chose a positive force to be pointing in the direction of the smaller arc-length. Since Eq.(1) is positive, we see that the planet which is displaced from the center is further pushed away from the center. Thus our planet and uniform-mass ring system is not stable, and the rings of Saturn cannot be uniform.

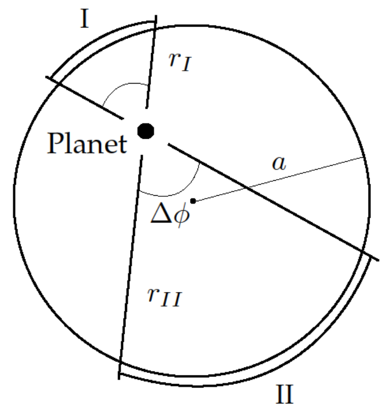

Figure 1: Planet of mass M_P is displaced from the center of a uniform massive ring of radius a.

2 A Definition

For the previous question, we completed the relevant calculations at two levels of rigor (three if we include the calculation in Appendix A), but we obtained the same answer in all cases. This is a standard quality of all calculations in physics: There is often a spectrum of rigor we can apply as we work toward a final answer.

At the highest level in this spectrum, one would answer the question with all relevant assumptions incorporated, with the correct mathematical framework, and with no approximations. Failing these conditions, a physics calculation becomes an estimate and moves in the direction of a back of the envelope calculation. A **back of the envelope** calculation is one that makes extensive use of simplifying assumptions and approximations to quickly answer a question. At the lowest level of rigor, one considers only the most crucial assumptions, employs non-rigorous mathematical arguments, and makes fanatical use of approximations like **order of magnitude** estimates (e.g., taking 6.022×10^{23} to be 10^{23}) and **scaling** (e.g., reducing Newton's law of gravitation to the law $F \sim 1/r^2$).

Since many important calculations in physics (e.g. elliptical orbits of our planetary system, thermal behavior of gases, electric fields between capacitor plates) fall short of the highest level of rigor, all such calculations are in some sense approximations. Thus we can argue that back of the envelope calculations do not exist starkly apart from standard calculations in physics, but instead, like these standard calculations, exists somewhere in a spectrum demarcated by logical rigor. Since the incorporation of mathematics into physics, physicists have worked to make their models of the world mathematically precise, but even such tightly constructed theories allow considerable space to derive correct physical results without deploying the full theoretical apparatus that physicists have built. This fact is just as relevant for formal calculations as it is for back of the envelope calculations.

Figure 2: Spectrum of rigor: Back of the envelope calculations, like standard physics calculations exist somewhere in a spectrum of rigor for physics calculations.

One important aspect of back of the envelope calculations is that they are distinguished from the calculations one might encounter in a physics class in that they are often open-ended and cannot be answered algorithmically. Thus there are multiple ways to complete the calculation, and the path you choose is greatly contingent on the weights you place on various assumptions and your chosen mathematical representation of the associated system. Not all such choices are equally valid, but there is still a large amount of room for divergent, yet mutually correct, approaches to a question.

Finally, we should note that the back of the envelope questions we are considering here do not concern the more colloquial "guesstimation" type questions outside of physics. Namely, we're not concerned with order of magnitude estimates of the kind used to determine the number of miles Americans drove last year. Such questions can be approached using techniques discussed in these notes, but they are not principally of concern to physics since they do not require us to consider physical principles in developing an answer.

3 General Setup for Calculations

Back of the envelope calculations usually have a few standard components. For clarity, we will write these components explicitly here, but note they are rarely ever implemented as cleanly as they are outlined below.

- **Physical foundations and principles:** Specify the physical principles and disciplines which are relevant to the calculation.
- **Mathematical analysis:** Use the rough mathematical forms of the physical principles and basic mathematics (i.e., calculus at most) to derive relationships between relevant variables. Dimensional analysis and limiting cases are useful for deriving results without an involved analysis, and scaling is useful for discerning the most important functional dependence in a final result.
- **Order of magnitude estimate:** Represent numerical quantities as orders of magnitude, namely as powers of 10, to simplify numerical calculations.

However, there is still one component which precedes all the others. Since it is not possible to solve a problem until it exists, the most important step in completing a back of the envelope calculation is defining the motivating question.

- **Define the question:** express the question in a way that it can be answered using the methods of physics.

As is often the case in physics, this is where much of the challenge comes from: recognizing that a problem even exists to be solved, and then further expressing the problem in a language so that is soluble. In these directions it is difficult to provide specific and universal guidance as how to define back of the envelope calculations, so in lieu of an algorithm, we provide examples. These examples begin with pre-processed problems in the sense that the motivating question already exists, even if it may not be precisely stated. Thus, these examples do not fully model the initial "question-articulating" process of back of the envelope calculations.

4 Examples

- **Using a pencil, you poke a hole in the bottom of a full water bottle (with its cap screwed off). How much time would it take all the water to drain out?**

  This is potentially a fluid mechanics problem, but the fact that the cap of the water bottle is screwed off implies the pressure above the water and outside the spout is the same. Thus we don't need to use Bernoulli's equation. Instead, by conservation of matter, we know that the height h of the water as a function of time should be related to the speed v at which water leaves the hole. For water of density ρ in a water bottle of cross sectional area A, the change in volume of water upon a height change of Δh

is $\rho A \Delta h$. If this decrease occurs in a time Δt, the amount of water which exits the hole in this time is $\rho \pi r^2 v \Delta t$, where r is the radius of the hole and v is again the speed of the water at the hole. Summing these contributions and noting they must together equal zero we have

$$\rho A \Delta h + \rho \pi r^2 v \Delta t = 0, \tag{2}$$

or the differential equation

$$\frac{dh}{dt} = -\frac{\pi r^2}{A} v. \tag{3}$$

By dimensional arguments, we can guess that the speed of the water exiting the hole is $v = \sqrt{2gh}$. We could also derive this more carefully beginning from Bernoulli's equations. With this expression for v, we can solve Eq.(3) to find

$$\begin{aligned} \int_{h_0}^{0} \frac{dh}{\sqrt{h}} &= -\frac{\pi r^2}{A} \int_0^{t_f} dt \sqrt{2g} \\ 2\sqrt{h_0} &= \frac{\pi r^2}{A} \sqrt{2g}\, t_f \end{aligned} \tag{4}$$

or

$$t_f = \left(\frac{R_b}{r}\right)^2 \sqrt{\frac{2h_0}{g}}, \tag{5}$$

where R_b is the radius of the circular cross section of the bottle. Eq.(5) is the time it takes the water starting from a height h_0 to completely drain from the open bottle. For a hole made by a pencil, we can take $r \approx 10^{-3}$ m. The height of a typical Poland Springs water bottle is about $h_0 \approx 15$ cm, and its radius is about $R_b \approx 2$ cm. With these values we find

$$t_f \approx 70 \text{ s}, \tag{6}$$

which is a reasonable time scale for this process. ■

- **By what percentage does your swing angle decrease as you ride on swing set? How many swings would it take for you to be roughly stationary?**

  Here, we are considering a person on a swing. In realistic swing systems, there are forces from air drag which oppose the motion of the swinging individual. There are also frictional forces from the point of contact between the chain and the supporting pole, but we will assume this is negiligible. To answer the above question requires us to consider energy and drag forces. Let's say you are a mass m object and you begin at an angle θ_0 from the vertical. The length of the chain's swing is ℓ. Therefore, your initial energy is $E_0 = mg\ell(1 - \cos\theta_0)$. After half a swing, let's say your angular apex is $\theta_0 - \Delta\theta$ where $\Delta\theta \ll \theta$. Thus your final energy is $E = mg\ell(1 - \cos(\theta_0 - \Delta\theta))$ and your change in energy is

$$\Delta E = mg\ell(1 - \cos(\theta_0 - \Delta\theta)) - mg\ell(1 - \cos\theta_0) \approx -mg\ell\Delta\theta \sin(\theta_0). \tag{7}$$

  We'll assume this loss in energy arises primarily from the force of air drag. Laminar flow which leads to a linear-in-velocity stokes drag is associated with slow moving objects. We will assume non-laminar flow. The force of air drag for non-laminar flow is $F = \frac{1}{2}\rho C_D A v^2$, where ρ is the density of air, C_D is the drag coefficient of the object, A is the cross sectional area of the object, and v is the speed of the object. We will approximate the speed during this first half-period as $v \simeq 2\theta_0 \ell/(T/2)$ where T is the period of the swing without drag. Thus the drag force on the object is

$$F \simeq 8\rho C_D A \frac{\theta_0^2 \ell^2}{T^2}. \tag{8}$$

The object moves a arc-length distance of $\Delta X = 2\theta_0\ell$ over the time we're considering. So the energy taken away from the object by drag forces is

$$\Delta E \simeq 16\rho C_D A \frac{\theta_0^3\ell^3}{T^2}. \tag{9}$$

Equating Eq.(7) and Eq.(9) and solving for $\Delta\theta$, we find

$$\Delta\theta \simeq \frac{16\rho C_D A\theta_0^3\ell^3}{T^2 mg\ell\sin(\theta_0)} = \frac{4\rho C_D A\theta_0^3\ell}{\pi^2 m\sin(\theta_0)}, \tag{10}$$

where we used $T = 2\pi\sqrt{\ell/g}$. The percentage decrease in the swing angle is then

$$\frac{\Delta\theta}{\theta_0} \simeq \frac{4\rho C_D A\theta_0^2\ell}{\pi^2 m\sin(\theta_0)}. \tag{11}$$

If you complete N full-period swings and finish at some final angle θ_f, then θ_0 relates to this final angle through

$$\theta_f = \theta_0\left(1 - \frac{\Delta\theta}{\theta_0}\right)^{2N}. \tag{12}$$

We include a power of 2 in this expression because $\Delta\theta/\theta_0$ corresponds to the fractional decrease from *half* a period of the swing. Thus, the number of swings it takes to reach the final angle is

$$N \simeq \frac{\ln(\theta_0/\theta_f)}{2\ln(1+\Delta\theta/\theta_0)}, \tag{13}$$

where we used $(1-x)^{-1} \simeq (1+x)$ for $|x| \ll 1$. Taking the swing to have a length $\ell \approx 2$ m, your mass to be $m \approx 70$ kg, your drag coefficient to be $C_D \approx 1$, your cross sectional area to be $A \approx 2$ m^2, and your initial angle to be $\theta_0 = \pi/4$, we find a fractional decrease of

$$\frac{\Delta\theta}{\theta_0} \approx 0.03, \tag{14}$$

and the number of swings it takes to reach the angle $\theta_f = 0.1$ rad (which we take to be the roughly stationary position), is

$$N \approx 40, \tag{15}$$

which encourages a live-testing for affirmation. ■

- **What is Young's modulus for a superball? (This assumes you've played with a superball before.)**

  A superball is a ball which conserves most of its energy upon collisions with hard surfaces. For example, if you were to drop a superball from a height of $h = 1$ m, the ball would hit the floor and rise back up to 1 m. Since superballs are defined by their energy conserving properties, this question requires us to equate two forms of energy: The elastic potential energy stored by the superball when it is slightly compressed on the floor, and the gravitational potential energy of the superball when it returns its initial height.

  If Y is Young's modulus of the superball, ΔL is the distance by which the superball is compressed after it hits the floor, L_0 is the length of the superball, and A is the area of contact between the compressed ball and the floor, then we have the elastic potential energy

$$U_{\text{elastic}} = \frac{1}{2}\left(\frac{YA}{L_0}\right)\Delta L^2. \tag{16}$$

For a ball of mass m starting from the height h, the gravitational potential energy is

$$U_{\text{grav}} = mgh. \tag{17}$$

Equating U_{grav} and U_{elastic} and solving for Y, we find

$$Y \simeq \frac{2mghL_0}{A\Delta L^2}. \tag{18}$$

A typical superball is about 1 pound or ≈ 0.5 kg. When it hits the floor after being dropped from a height $h = 1$ m, we could estimate that it compresses by $\Delta L \approx 1$ cm, and that its area of contact with the floor is $A \approx \pi(1\text{ cm})^2$. Using these values, we find

$$Y \approx 1.2 \times 10^7 \text{ kg/m}\cdot \text{s}^2, \tag{19}$$

or $\sim 10^7$ Pa. This matches the standard Young's modulus of rubber (of which all superballs are made) of 0.01–0.1 GPa. (http://www.bestech.com.au/wp-content/uploads/Modulus-of-Elasticity.pdf). ■

5 Utility

We presented back of the envelope calculations as a way to quickly find an answer to a physics problem without having to wade through the complicated waters of a more precise analysis. But beyond their solubility, such calculations are useful as ways to survey the territory before an extended investigation. When starting a new project, it is not always clear at the start whether the answer one may find in the end is physically relevant or even important, and thus whether said answer justifies all the preceding effort. With a back of the envelope calculation, one can make a quick order-of-magnitude estimate of the result and can thus obtain a quantitative sense of the calculation's physical relevance ever before the main analysis. Thus because of where they exist in the spectrum of rigor, back of the envelope calculations are useful as precursors to more precise calculations.

Limits of Utility: One thing back of the envelope calculations are decidedly inadequate for is proving theorems. For example, you might use such calculations to estimate the binding energy of the hydrogen molecule ion, but you could not perform a back of the envelope calculation to prove the validity of the variational method. In general, any investigation which requires you to delve into the internal mathematical structure of a theory cannot be probed with calculations which use approximations and heuristics to circumvent that structure.

6 References for practice

Although lessons in back of the envelope calculations are typically not included in the standard subject texts of physics, there are several books which are solely devoted to providing practice in these kinds of calculations. A few such texts are listed below.

– ***Thinking like a physicist* by N. Thompson:** Contains more than a hundred problems each of which encourage a flexible application of standard principles and models in physics.

– ***Physics of superheroes* by J. Kakalios:** Using rough approximations and hardly any mathematics beyond algebra, this text discusses many of the ways the superheroes of American lore violate or don't violate the laws of physics.

– ***Art of Insight in Science and Engineering* by Sanjoy Mahajan:** A philosophical and yet practical book which discusses the general motifs for representing and applying science and engineering knowledge. Includes end of chapter problems

– **Purcell's "Back of the envelope" calculations:** Not a textbook but an online resource containing many back of the envelope questions and solutions. (http://ajp.dickinson.edu/Readers/backEnv.html)

A Planet-Ring Instability: Rigorous Approach

In this section, we derive the results of the inciting example more rigorously using a Newtonian picture of the planet-ring dynamics. Although we don't consider it here, it is also possible to obtain these results from the Lagrangian formalism.

As in the main text we will calculate the force on the planet in order to determine the planet's direction of motion after it has been perturbed from the center of the ring. The situation is depicted in Fig. 3. We place our coordinate system at the ring's center and assume the planet remains in the plane of the ring.

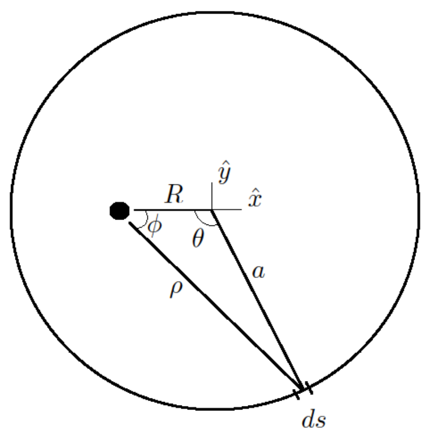

Figure 3: Planet-ring system with coordinate variables. ρ is the distance from the planet to the differential length ds; R is the distance from the planet to the ring's center; a is the distance from the ring's center to the differential length ds; ϕ is the angle between sides R and ρ; θ is the angle between sides R and a. We will take $R \ll a$.

By Newton's Law of Gravitation, the net force exerted on the planet by the ring is then

$$\vec{F} = GM_p \int_0^{2\pi a} ds\, \lambda_0(s)\, \frac{\hat{\rho}}{\rho^2}, \tag{20}$$

where s is the arc-length position parameter of the ring, and $\hat{\rho}$ is a unit vector pointing from the planet towards the arc-length element ds. We wrote the mass density $\lambda_0(s)$ as a function of s for full generality, but for this system the mass density is constant. Using the geometry implied by our coordinate system, we can covert this force into a more useful form. First, from the law of cosines, we have

$$\rho^2 = R^2 + a^2 - 2aR\cos\theta. \tag{21}$$

From the figure, we see that the unit vector $\hat{\rho}$ is

$$\hat{\rho} = \cos\phi\,\hat{x} - \sin\phi\,\hat{y}. \tag{22}$$

Forming a right triangle with an acute angle at ϕ, the other acute angle at ds, and the right angle at a specific point to the right of the circle's center, we can show that

$$\rho\cos\phi = R - a\cos\theta \tag{23}$$
$$\rho\sin\phi = a\sin\theta, \tag{24}$$

so that Eq.(22) becomes

$$\hat{\rho} = \frac{1}{\rho}\Big[(R - a\cos\theta)\hat{x} - a\sin\theta\hat{y}\Big] \tag{25}$$

Plugging Eq.(25) into Eq.(20) and changing integration variables $ds = ad\theta$, we see (since it consists of the integration of an odd function over an even domain) there is no force in the $\hat{y}$ direction. The resulting force is then

$$\begin{aligned}\vec{F} &= GM_p\int_0^{2\pi} ad\theta\,\lambda(\theta)\,\frac{(R-a\cos\theta)}{(R^2+a^2-2aR\cos\theta)^{3/2}}\hat{x}\\ &= 2GM_p a\int_0^{\pi} d\theta\,\lambda(\theta)\,\frac{(R-a\cos\theta)}{(R^2+a^2-2aR\cos\theta)^{3/2}}\hat{x}\end{aligned} \tag{26}$$

where we took $\lambda(\theta) \equiv \lambda_0(s = a\theta)$, and we used the properties of the cosine function to cut the domain of integration by two. It will prove analytically wise to make the above integral dimensionless. Doing so we have

$$\vec{F} = \frac{2GM_p}{a}\,\frac{I_\delta[\lambda]}{(1+\delta^2)^{3/2}}\,\hat{x}, \tag{27}$$

where

$$I_\delta[\lambda] \equiv \int_0^{\pi} d\theta\,\lambda(\theta)\frac{\delta - \cos\theta}{\big(1 - 2\delta\cos\theta/(1+\delta^2)\big)^{3/2}}, \tag{28}$$

and $\delta \equiv R/a$. We write $I_\delta[\lambda]$ to indicate that while I is a functional of λ it is a function of δ. We note that Eq.(27) is zero if $\delta = 0$, that is, if the planet is at the center of the ring. We are interested in the situation where the planet is only slightly perturbed from its center, i.e., where $\delta \ll 1$. So, to investigate the stability of the system we should expand Eq.(27) and Eq.(28) as power series in δ. For I we have

$$I_\delta[\lambda] = \int_0^{\pi} d\theta\,\lambda(\theta)\Big(-\cos\theta + \delta(1 - 3\cos^2\theta) + 3\delta^2(\cos^2\theta - \tfrac{5}{2}\cos^3\theta)\Big) + \mathcal{O}(\delta^3). \tag{29}$$

This result is general for all $\lambda(\theta)$ and hence could be used to investigate the orbital stability of various mass distributions. For the case of a constant mass density, we have

$$\begin{aligned}I_\delta[\lambda] &= \lambda_0\delta\int_0^{\pi} d\theta\,(1 - 3\cos^2\theta) + \mathcal{O}(\delta^2)\\ &= -\frac{\pi\lambda_0\delta}{2} + \mathcal{O}(\delta^2).\end{aligned} \tag{30}$$

We note that the first term in Eq.(29) integrates to zero for a constant mass density. By Eq.(27), we then have

$$\vec{F} = -\frac{\pi GM_p\lambda_0}{a}\left(\frac{R}{a}\right)\hat{x} + \mathcal{O}(\delta^2). \tag{31}$$

From Fig. 3, we recall that this force was computed by first assuming an initial displacement of the mass in the negative $\hat{x}$ direction. The sign of Eq.(31) indicates that the force of the planet is in the direction of this initial perturbation, and we can therefore conclude that the planet-(uniform)-mass ring system is unstable.