A Compact and Deliberately Incomplete Introduction to GFD

Robert Friel

March 18, 2012

Contents

Chapter 1

The GFD Bestiary

(***NOTES: pedlosky [p. 48] and vallis [p. 66] disagree on whether the traditional approx. is a small H/L limit or a small H/a limit. we want the section on "proposition 1" to cover only small H/a limits; need to determine whether or not the traditional approximation should go here. what vallis calls the "shallow fluid approx" should definitely go there, but surely there's a less stupid name for it out there?)

1.1 Preface

Geophysical Fluid Dynamics (GFD) is the study of a set of approximations that confer analytic insight into the motions of the atmosphere and ocean – analytic insight that would not be available from the un-approximated Navier-Stokes equations alone. Because there are many different types of approximations in GFD, and because several of them are often applied at the same time, it it common for people learning the subject for the first time to get lost in a storm of buzzwords. Hence, before launching into a detailed description of any of these approximations, I want to present an overview that catalogues them and explains how they relate to one another. That is the purpose of this chapter.

GFD is the study of two fluids (the atmosphere and the ocean) that form a thin layer of varying density on the surface of a large, roughly spherical planet that is rotating. Let's break that down. The objects of study in GFD are fluids which

1. exist on the surface of a *rotating sphere*, and are *influenced significantly* by the sphere's rotation

2. typically execute much larger and faster motions in the plane *parallel* to the sphere's surface ("horizontally") than they do in the direction *perpendicular* to it ("vertically")

3. have a vertical extent that is much *smaller* than the sphere's radius

4. have *density variations* in the vertical direction that are well-described by certain simplified models

These four properties (which we will call "Properties 1 to 4") each form the basis for their own set of approximations:

Property 1 says that fluid motions in GFD are affected by the planet's rotation. The degree to which a given fluid motion is affected by the planet's rotation is quantified by the *Rossby number* Ro and the Ekman number Ek (smaller Rossby and Ekman numbers means rotation is more influential). Roughly speaking, the Rossby number quantifies how strong the rotational effects are relative to advective effects, and the Ekman number quantifies how strong the rotational effects are relative to viscous effects.

In GFD, we typically *assume* that the Ekman number is small enough that viscosity is approximately irrelevant (technically, confined to a small boundary layer – more on this later). If, in addition, the Rossby number of a motion is small, we can use something called the *quasi-geostrophic approximation*.

Property 2 is called "having a small aspect ratio," and the extent to which a motion possesses it is quantified by a number called the *aspect ratio*, δ. The approximation made possible by Property 2 is called the *hydrostatic approxmation*. If we additionally add the constraint that the fluid is incompressible, the hydrostatic approximation becomes the *shallow water approxmation*.

Property 3 is the justification for an approximation which, for some reason, is called the *shallow-fluid approximation*. (Confusingly, this has nothing to do with the shallow *water* approximation mentioned in the previous paragraph, which relies on Property 2, not Property 3.)

Property 4 is a statement about ways that the density of the atmosphere and ocean vary in the vertical direction. In GFD jargon, such variations are called *stratification*. Our knowledge about the stratification of the ocean and atmosphere lead to two approximations that describe fluid motions as small deviations away from an unmoving background state. These are called the *Boussinesq approximation* (typically used for the ocean) and the *anelastic approxmation* (typically used for the atmosphere).

A few remarks before we continue. First, although you probably know that planets aren't perfect spheres, we won't be interested in deviations from sphericity here. The planets we'll consider will be treated as perfect spheres. (And I'll use the word "earth" instead of "planet" from now on, although GFD can be used to study other planets.) Second, an important thing to note about GFD is that its basic principles apply equally well to the earth's atmosphere and ocean, even though those two fluids might appear very different at first glance. The differences between the two will get discussed a little later. Roughly, you can think of the atmosphere as an ideal gas (i.e., a compressible fluid) whose density varies a lot in the vertical direction, and the ocean as a roughly incompressible fluid (whose density, thus, doesn't vary that much). Except for this important difference, our four properties and their consequences apply equally well to both, and can be studied

in themselves without reference to either fluid in particular.

1.2 Untamed reality: the Navier-Stokes equations

1.2.1 Navier-Stokes review

I'm going to assume that you have studied basic fluid mechanics, and that you know the Navier-Stokes equations, the fundamental equations of fluid flow. As a refresher, here they are in coordinate-independent form:

$$\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{v} + \nabla\Phi \tag{1.1}$$

$$\frac{D\rho}{Dt} = -\rho(\nabla \cdot \mathbf{v}) \tag{1.2}$$

These describe the evolution in time of a vector field $\mathbf{v}$ in three-dimensional space whose value at a point $\mathbf{x}$ at time t gives the instantaneous velocity of the fluid parcel passing through $\mathbf{x}$ at t. D/Dt is the material derivative, defined as

$$\frac{D\mathbf{f}}{Dt} = \frac{\partial\mathbf{f}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{f} \tag{1.3}$$

for an arbitrary vector field $\mathbf{f}$.

The equation (1.1) is called the momentum equation, and says the the acceleration of a fluid parcel (the left-hand side) equals the sum of the forces on it, divided by its mass (the right-hand side). The forces are a pressure force (first term), viscous dissipation (second term), and conservative forces like gravity (third term, where Φ is the potential for these forces).

Equation (1.2), called the continuity equation (or mass conservation equation), says that since mass is never created or destroyed, a parcel can change density only when a divergence in velocity causes its sides to expand or contract. If we assume that our fluid is *incompressible*, meaning that parcels never change density, then $D\rho/Dt = 0$, and so (1.2) says that

$$\nabla \cdot \mathbf{v} = 0 \tag{1.4}$$

If we take the curl of the momentum equation (1.1), we we get an equation for the vorticity $\omega = \nabla\times\mathbf{v}$, one in which the curl has killed off the ∇p term. If we also have incompressibility (1.4), then we know the divergence and curl of our velocity field, which is often enough information to make a solution unique. So given incompressibility, we don't usually need to know about the evolution of p itself.

On the other hand, for a compressible fluid, eqs. (1.1)-(1.2) are not enough, and we need an additional equation describing how p is determined, called an "equation of state." Typically this will come from thermodynamic considerations, and thus will involve the

temperature T, which will need an evolution equation of its own. The two equations that close the system will, in very generic form, look something like this:

$$p = p(\rho, T) \tag{1.5}$$

$$\frac{DT}{Dt} = \dot{T} \tag{1.6}$$

These just say that the pressure is some known function of ρ and T, and that the material derivative of T is given by some known function $\dot{T}$ (here, I've left it completely unspecified). One familiar equation of state, which applies well to the atmosphere, is the ideal gas law $p = \rho R T$. The equation of state for ocean water is more complicated.

1.2.2   Navier-Stokes and GFD

The atmosphere and ocean, like everything else on the earth (including us), move with the earth as it turns. The motions we see as weather, ocean currents, etc. are motions *relative to* this basic solid-body rotation that all terrestrial beings share. Since we are interested in these relative motions, we will move from a fixed, inertial coordinate frame to one that follows the earth's rotation.

I will not go through the details of transforming from an inertial frame to a rotation frame. (This process is described in any textbook on classical mechanics or GFD.) Basically, what happens is the following. For fairly intuitive reasons, scalars and spatial gradients are the same in both frames, which means that the continuity equation (1.2) and the right-hand side of the momentum equation (1.1) remain the same. On the other hand, time derivatives of vectors pick up a term from the rotation, which affects both $\mathbf{u}$ (the time derivative of parcel position) and $D\mathbf{u}/Dt$ (the time derivative [following the parcel] of parcel velocity). The result is that, for a coordinate system with angular velocity vector $\mathbf{\Omega}$ (assumed constant),

$$\left(\frac{D\mathbf{v}_I}{D\mathbf{t}}\right)_I = \left(\frac{D\mathbf{v}_R}{D\mathbf{t}}\right)_R + 2\mathbf{\Omega} \times \mathbf{v}_R + \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r}) \tag{1.7}$$

where $\mathbf{v}_I$ is the velocity measured in the inertial frame, $\mathbf{v}_R$ is the velocity measured in the rotating frame, and the I and R subscripts on the derivatives indicate derivatives measured in the inertial (I) and rotating (R) frames.

The second term on the right-hand side of (1.7) is called the *Coriolis acceleration*, and is the *sole source* of all the effects of the earth's rotation in GFD. When we talk about "effects of the earth's rotation," we mean "effects of the Coriolis acceleration."

The third term, called the *centrifugal acceleration*, can be written as the gradient of a potential, and hence can be bundled into the potential Φ in the momentum equation. It is usually unimportant for GFD, and is typically just ignored.

Thus, combining (1.7) (without the third term) with the inertial-frame momentum equation (1.1), we obtain the rotating-frame momentum equation (the subscript R has

been dropped):

$$\frac{D\mathbf{v}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{v} + \nabla\Phi \tag{1.8}$$

Note that if we moved the $2\boldsymbol{\Omega} \times \mathbf{v}$ term to the other side, it would look like just another force on the parcel. That is, dynamics in a rotating frame looks like dynamics in an inertial frame with an additional force (a "fictitious force").

The Navier-Stokes equations in a rotating frame are (1.8) plus (1.2) and either incompressibility or an equation of state. These are the fundamental, exact equations for fluid motions on a rotating planet. In principle, they can describe any motion that happens in the atmosphere and ocean, from the smallest space and time scales to the largest. But because of their nonlinearity, these equations are difficult to solve exactly. And the fact that they encompass all space and time scales is a bug rather than a feature when it comes to gaining intuition: the equations do not tell us what any individual type of motion looks like independent of all others. This is why the field of GFD exists.

The remaining sections in this chapter cover a series of approximations to the rotating N-S equations which add intuition and practical usability. The purpose of this chapter is to give a sense of what the important approximations are and how they fit together. Thus, I will describe these approximations in broad strokes rather than rigorously defining and deriving them.

1.3 How to tame reality: scaling arguments and balances

1.3.1 Scaling primer

This section deals with two of the most important approximations in GFD, *geostrophic balance* and *hydrostatic balance*. These approximations exploit Propositions 1 (rotation is important) and 2 (small aspect ratio), respectively. The section on geostrophic balance will also cover the argument that is used to justify ignoring viscosity in GFD, since this argument also depends on the importance of rotation, albeit in a different sense.

The type of argument used to justify these two approximations is called a *scaling argument*. You may not be familiar with this type of argument, and I will use the first part of this section to explain how it works.

A scaling argument looks sort of like a crossbreed between perturbation theory and dimensional analysis, but it is distinct from both of those topics. At this point, we are going to need to use some basic regular perturbation theory. I am going to assume that you know how regular perturbation theory works, but just in case, here is a brief summary.

Perturbation theory

Suppose we have a math problem that can be stated in this form:

$$L_0 u = 0 \tag{1.9}$$

where L_0 is some sort of operator and u is a unknown we want to solve for. In the case that interests us, L_0 is a differential operator and u is a function, so this is a differential equation.

Suppose we know how to solve (1.9) to find u. But now imagine that we are interested instead in another, harder problem, of the form

$$L_0 u + \epsilon L_1 u = 0 \tag{1.10}$$

Here ϵ is a parameter that we will assume is small (in a sense to be explained soon). L_1 is some operator that makes the problem hard to solve analytically. We call this the "perturbed problem"; (1.9) is the "unperturbed problem." What we want to do is form an approximate expression for u that converges toward perfect accuracy as ϵ tends to zero. This will allow us to approximately solve the perturbed problems for small perturbations away from the unperturbed problem, even if we can't solve the perturbed problem exactly.

Note that (1.10) defines a family of problems indexed by ϵ (each value of ϵ, when plugged into (1.10), gives us a different problem). We will write $u(\epsilon)$ for the solution to the version of (1.10) corresponding to the argument ϵ. (For instance, $u(0)$ is just the solution to (1.9).)

Now, we make the assumption that $u(\epsilon)$ is an analytic function near $\epsilon = 0$. This means that we can expand it as

$$u(\epsilon) = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \dots \tag{1.11}$$

for entities (in our case, functions) u_0, u_1, etc. which are not themselves functions of ϵ. Now plug this back into (1.10). Grouping together terms in powers of ϵ, we get

$$(L_0 u_0) + \epsilon(L_0 u_1 + L_1 u_0) + \epsilon^2(L_0 u_2 + L_1 u_1) + \dots = 0 \tag{1.12}$$

By hypothesis, $u(\epsilon)$ solves (1.10) for every value of ϵ. Thus, (1.12) must hold for every value of ϵ. The unique power series in ϵ that equals the zero function (the right-hand side of (1.12)) is the power series all of whose coefficients are zero. Thus we can conclude that every *individual* term in (1.12) is zero, i.e.,

$$L_0 u_0 = 0 \tag{1.13}$$
$$L_0 u_1 + L_1 u_0 = 0 \tag{1.14}$$
$$L_0 u_2 + L_1 u_1 = 0 \tag{1.15}$$
$$\dots \tag{1.16}$$

The first of these equations just says that u_0 solves the unperturbed problem, as expected. Plugging this solution for u_0 into the second equation gets rid of the nasty operator L_1 (since we're just applying it to a known function), leaving just an inhomogeneous version of the unperturbed problem. Assuming we know how to solve *that*, we can just keep

going down the chain, solving for as many u_i functions as we choose to. In the end, we will typically just keep u_0 and u_1. The goal is not to get an extremely accurate approximation for any value of ϵ, which will generally be impossible; the goal is to get an approximation that works when ϵ is so small that all but the first term (or the first few terms) are very small.

It will be worth elaborating on that a bit. It's easy to get oneself confused when doing perturbation theory, because the entire procedure depends on the idea that ϵ is "small," so one is constantly saying things like "for small ϵ" and "for ϵ small enough" – yet the procedure itself doesn't give you any information about *how small* is "small enough."

The problem is that each term in the series has a constant in front of it – u_0, u_1 and so on – and these can be arbitrarily large. (Note: I call them constants, but of course they could be functions or something; they're "constants" in that they don't depend on ϵ.) The nice thing about a perturbation series is that the ratio of each term to the previous one is ϵ, so that as we make ϵ small, the later terms grow more and more insignificant compared to the earlier ones. But all this gets us is an unhelpful assurance that *in the limit* as $\epsilon \to 0$, the series will converge towards being a perfect approximation. We don't know *a priori* how small it has to be. In order to know for sure, we'd have to know every single constant in the infinite perturbation expansion (because before we check, how are we to know that, say, u_{151} isn't ten-to-the-power-of-some-unimaginably-huge-number?).

In practice, a person who does perturbation theory will simply take a leap of faith and say, "I am going to guess that, for the kinds of ϵ values I encounter in practice, everything beyond the first few terms of this series will be negligible." They calculate the first few terms, check the results against reality, and if the approximation makes good predictions, they declare the effort a success. One must always keep in mind that one does not really know what "a small enough ϵ" means in any given problem. Instead, one uses whatever ϵ one is given by the physical parameters of the problem, and hopes that it turns out to be small enough.

Scaling

Scaling arguments are formally similar to perturbation theory arguments. The difference is that in a scaling argument, we do not consider small deviations from an "unperturbed problem," but instead *cause* certain terms in a problem to *become* small by pulling out factors that reflect estimates of the size of each term in the problem. Typically there will be no identifiable "basic" state relative to which "small deviations" are occurring.

The following simple example captures the essence of the idea.

Suppose that I am trying to solve Laplace's equation in two dimensions.

$$\frac{\partial^2 V(x,y)}{\partial x^2} + \frac{\partial^2 V(x,y)}{\partial y^2} = 0 \tag{1.17}$$

(For concreteness, imagine that I'm solving for the electric potential $V(x,y)$ in some 2D do-

main.) In a typical problem of this sort, one would be presented with boundary conditions, and would then solve for the behavior on the interior through something like separation of variables. But suppose that, rather than getting information about what's happening at the boundary, I get some information about what the solution on the interior *looks like* – information that constrains the solution without exactly determining it.

Specifically, suppose that $V(x,y)$ is characterized by a typical *horizontal length scale* X and a typical *vertical length scale* Y. Suppose that, in particular, I know that the ratio X/Y, which I will suggestively call ϵ, is much less than one. (We're going to use ϵ in a perturbation expansion, so "much less than one" here means the same thing as "small enough" in perturbation theory – see the discussion above.)

What on earth do I mean by "length scales"? The actual *meaning* of this term can be a bit subtle, and I will go over it later, after I show how the procedure actually works.

To begin the scaling argument, we introduce new variables $\tilde{x}$ and $\tilde{y}$ such that $x = X\tilde{x}$ and $y = Y\tilde{y}$. That is, the new variables – called *nondimensional variables* – are the old variables with the relevant scales factored out. Now, we want to recast the whole problem in terms of these new variables. So we will rewrite things in terms of a new function of the new variables, $\tilde{V}(\tilde{x}, \tilde{y})$ (where of course $\tilde{V}(\tilde{x}, \tilde{y}) = V(X\tilde{x}, Y\tilde{y}) = V(x,y)$). The chain rule tells us that

$$\frac{\partial}{\partial x} = \frac{1}{X} \frac{\partial}{\partial \tilde{x}} \tag{1.18}$$

and likewise for y. Thus (1.17) turns into

$$\frac{1}{X^2} \frac{\partial^2 \tilde{V}(\tilde{x}, \tilde{y})}{\partial \tilde{x}^2} + \frac{1}{Y^2} \frac{\partial^2 \tilde{V}(\tilde{x}, \tilde{y})}{\partial \tilde{y}^2} = 0 \tag{1.19}$$

or

$$\frac{\partial^2 \tilde{V}(\tilde{x}, \tilde{y})}{\partial \tilde{x}^2} + \epsilon^2 \left(\frac{\partial^2 \tilde{V}(\tilde{x}, \tilde{y})}{\partial \tilde{y}^2} \right) = 0 \tag{1.20}$$

recalling that $\epsilon = X/Y$. Hey, look, we have a problem with a small ϵ in it! Time to write a perturbation series:

$$\tilde{V}(\tilde{x}, \tilde{y}) = \tilde{V}_0(\tilde{x}, \tilde{y}) + \epsilon \tilde{V}_1(\tilde{x}, \tilde{y}) + \epsilon^2 \tilde{V}_2(\tilde{x}, \tilde{y}) + \ldots \tag{1.21}$$

"Not so fast!" exclaims the mathematically savvy reader. "The small parameter multiplies the largest derivative in the variable $\tilde{y}$, which means that this is probably a *singular perturbation problem*, which means that the *regular* perturbation theory you described in the previous section shouldn't be able to solve it!" Right you are, Mathematically Savvy Reader. And yet we're going to go right ahead with it. As it turns out, this actually works, but I won't be able to explain until a little while later. (The same explanation will, hopefully, also clear things up for people to whom this paragraph makes no sense whatsoever.)

Anyway, if we plug our perturbation series into (1.20), it's not hard to see that the equation governing the lowest order term $\tilde{V}_0$ is just

$$\frac{\partial^2 \tilde{V}_0(\tilde{x}, \tilde{y})}{\partial \tilde{x}^2} = 0 \tag{1.22}$$

This would have been true even if the $\tilde{y}$-derivative had had just an ϵ rather than an ϵ^2 in front of it – the point is just that any term with a small parameter in front of it is not going to enter into the lowest-order term of a perturbation expansion.

Typically, when doing a scaling argument, we don't even go to the trouble of writing out a perturbation expansion. We just use the appearance of a small parameter in front of a term as a justification for *deleting* that term, which is really a shorthand for "if we made a perturbation expansion, and decided to use only the lowest-order term, the equation it would satisfy is the one we get when we delete any term with a small parameter in front of it."

Now we have an equation for the function $\tilde{V}$. (We'll omit the subscript 0, since in a scaling argument we only ever care about the lowest order.) But what we really wanted was the function V. Of course, going from one to another is easy. Just solve (1.22) for $\tilde{V}$, then write $V(x, y) = \tilde{V}(x/X, y/Y)$. That's our solution – to lowest order in ϵ, anyway.

(At this point the Observant Reader, after exchanging a frustrated glance with the Mathematically Savvy Reader, raises her voice to ask just how I plan to solve (1.22) for a function of $\tilde{x}$ and $\tilde{y}$ when that equation doesn't even have $\tilde{y}$ in it, and thus doesn't constrain the $\tilde{y}$-dependence of $\tilde{V}$ at all. The answer is "I can't." More on this point in a moment.)

So the scaling procedure goes like this:

1. identify scales for relevant variables (these can be spatial variables, as in this problem, or they can be any other sort of variable)

2. define non-dimensional variables

3. write out the problem in terms of the non-dimensional variables

4. delete any term with a small parameter in front of it

5. solve the resulting problem for the nondimensional solution

6. write out the dimensional solution in terms of the nondimensional solution and the scales

But what does this procedure actually do, and how do we pick the scales? The justification for the procedure starts with the idea that the values of the variables in the problem tend to remain close to the value we choose for the scale, so that the non-dimensional variables will tend to remain around 1. In our case, that would mean that, e.g., the derivative $\partial^2 V/\partial x^2$ will really have a numerical value close to $1/X^2$ at every point in the domain, and so $\partial^2 \tilde{V}/\partial \tilde{x}^2$ will have a numerical value close to 1 at every point in the domain.

This is what justifies the perturbation approach, because if all the non-dimensional terms have around the same value by themselves, then any one of them with an ϵ in front of it will really be smaller than the others. If we didn't choose the scales right, we wouldn't have a real perturbation problem. We could try to form a perturbation series, but the higher-order terms would have big constants in front of them (think about why), and thus the lowest-order term would not reflect reality well.

Note that the choice of scale is an assumption about what the *solution* will look like. A single equation may allow many different types of behavior; choosing scales is like saying "I want to look at solutions that have a certain set of properties." In our case, we chose to look at solutions that vary much more slowly in the y direction than in the x direction. If this kind of behavior is possible in the system we're studying, then the scaling analysis will give us equations appropriate for it. The procedure does not give us a comprehensive list of the behaviors that can emerge from a given equation, or tell us why one behavior will happen rather than another. Instead, it tells us what a given type of behavior will look like, if it occurs.

This consideration provides the answer to the Mathematically Savvy Reader's question from earlier. What the MSR was worried about is that the differential equation (1.20) might have some solutions for which the second $\tilde{y}$-derivative gets arbitrarily large compared to the $\tilde{x}$-derivative (at some points), so that even the ϵ^2 doesn't assure that the second term remains small compared to the first. These solutions can't be represented as zeroth-order solutions plus small corrections, since these behaviors don't exist at all in the unperturbed case (where no matter how large the $\tilde{y}$-derivative gets, it simply doesn't affect the equation at all). This is called a singular perturbation problem, as opposed to regular perturbation problems, which are the sort I described earlier. If the small parameter multiplies the highest-order derivative (in any particular variable) in an equation, that's a sign that a problem is likely to be singular. In singular problems, the behavior of the perturbed problem doesn't converge to the behavior of the unperturbed problem even as $\epsilon \to 0$ (because of the special behaviors that appear for any finite ϵ, but not for $\epsilon = 0$), so the problem needs to be modified before infinite series methods can be applied.

There are well-understood techniques for solving such problems, but in this case we don't need to use them, because we're okay with excluding the special solutions for which the nondimensional derivative gets arbitrarily large and counteracts the smallness of ϵ. In fact, we *want* to ignore those solutions. When we make a choice of scale, we we assume that the nondimensional variables will have values close to one. That's what we *mean* when we say things like "the horizontal scale is 100 meters" (or whatever). That's a description of the sorts of solutions we're interested in. If there exists a solution to our nondimensional equations where the nondimensional quantities aren't all close to one, that solution doesn't really have the scales we said it had at the outset. When we made a choice of scale, we declared which sorts of solutions we were interested in, and there's no reason to bend over backward in order to find solutions that don't actually fit that description. So, no point in trying to find the singular behaviors. We're not interested in them anyway.

If ϵ multiplies the highest *time*-derivative in an equation, the problem will typically exhibit *multiple time-scales*. What that means is that we'll see really quickly oscillating solutions that don't appear in the unperturbed problem. (Every time you take a derivative of a sinusoid with frequency ω, you get another multiplicative factor of ω, so if ω is as big as $1/\epsilon$, the term with an ϵ won't be small.) On the other hand, if the high derivative with the ϵ is a *spatial* derivative, the problem has a *boundary layer*: a small region, typically near the boundary, where spatial variation is large. Both of these behaviors will come up later in our investigations.

Scaling arguments are, basically, a way of exploiting the characteristics of a desired solution to create a perturbation problem. There is an importance difference, however, between a scaling argument and a more standard perturbation argument. In a standard perturbation argument, we begin with a solvable problem, the "unperturbed problem," whose solution is "the zeroth order solution." We then add a small change to the unperturbed problem, and solve to find the approximate change to the zeroth order solution caused by the correction. The whole point is to get *beyond* the zeroth-order solution, to observe how the perturbation changes it. On the other hand, in a scaling problem, typically we retain the zeroth order solution *and nothing else*. The solution is so crude that it actually does not contain any information about how the small parameter differs from zero!

For instance, in the example above, we are effectively acting as though X/Y is *zero*, not just small, since in (1.22) we have removed all dependence on X/Y. However, this approximation is *not* made when we convert the nondimensional solution into the dimensional solution. We solve the *nondimensional* problem as though X/Y were zero, then use the result in a context where X/Y is not zero, just small.

Balances

The notion of a "balance" is really nothing new, but I figure I should explicitly introduce the term before using it.

1.4   How to tame reality: linearization