Chapter 1

What is a singular perturbation?

Mathematical equations arising from physical sciences contain parameters. Perturbation theory examines parameter dependence of solutions locally. To present basic ideas simply, consider a one-parameter family of functions: For each $x$ in a set $R$ and real parameter $\epsilon$ in a punctured neighborhood of $\epsilon = 0$, the values of the functions $f(x, \epsilon)$ are in a metric space. The range is a metric space so that convergence of $f$’s as $\epsilon \to 0$ can be discussed. $f(x, \epsilon)$ is to be regarded as a solution of some set of equations containing $\epsilon$ as a parameter.

The equations are called a regularly perturbed problem if all solutions $f(x, \epsilon)$ converge uniformly on $R$ as $\epsilon \to 0$. If there is a solution which does not converge uniformly, the problem is called singularly perturbed. Notice that the category, regular or singular, is formulated in terms of the solutions, and not the equations.

This abstract definition of singular perturbation is very broad. But practical problems draw attention to a few dominant categories of singular behavior. What follows is a mini-survey of

Prototypical examples

Singularly perturbed polynomial equations

For $\epsilon > 0$, the polynomial equation

$$\epsilon z^8 - z^3 - 1 = 0$$

(1.1)

has 8 complex roots. In the language of the preceding general discussion, the set $R$ is the integers $1, 2, \ldots 8$ labeling the roots, and the metric space of the
roots \( z = f(k, \epsilon) \), \( k = 1, \ldots, 8 \), is the complex numbers. Figure 1.1 displays numerical approximations of the 8 roots for the sequence of \( \epsilon \)'s, \( \epsilon = 2^{-k} \), \( k = 1, \ldots, 10 \). Three of the roots appear to be converging to the cube roots of \(-1\),

\[ (-1)^{\frac{1}{3}} = e^{\frac{i\pi}{3}}, e^{i\pi}, e^{\frac{2i\pi}{3}}. \]

as \( \epsilon \to 0 \). This is easy to see: Setting \( \epsilon = 0 \) in (1.1) gives the reduced equation

\[ z^3 + 1 = 0. \]  

Figure 1.1: The roots (dots) and the approximated roots (circles)

The remaining 5 roots are diverging: Figure 1.2 is a log-log plot of the positive, real root as a function of \( \epsilon \). It appears that the diverging roots scale with \( \epsilon \) like \( \epsilon^{-\frac{1}{5}} \). If the polynomial equation (1.1) were regularly perturbed, all 8 roots would converge as \( \epsilon \to 0 \). Hence, it is singularly perturbed.
Chapter 1. What is a singular perturbation?

If it is assumed that the roots of (1.1) exhibit algebraic scalings with $\epsilon$ as $\epsilon \to 0^+$, direct constructive approximation is easy. Roots that scale like $\epsilon^{-p}$ as $\epsilon \to 0$ can be represented as

$$z(\epsilon) = \epsilon^{-p} Z(\epsilon)$$  \hspace{1cm} (1.3)$$

where it is assumed that $Z(\epsilon)$ converges to a non-zero value as $\epsilon \to 0$. The scaled version of (1.1) in terms of $Z$ is

$$\epsilon^{1-8p} Z^8 - \epsilon^{-3p} Z^3 - 1 = 0.$$  \hspace{1cm} (1.4)$$

The limit $\epsilon \to 0^+$ is examined. At this point, the exponent $p$ is undetermined,
so the limit equation depends on $p$, as shown in the table below:

<table>
<thead>
<tr>
<th>$p$</th>
<th>limit equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p &lt; 0$</td>
<td>$-1 = 0$</td>
</tr>
<tr>
<td>$* p = 0$</td>
<td>$Z^3 + 1 = 0$</td>
</tr>
<tr>
<td>$0 &lt; p &lt; \frac{1}{5}$</td>
<td>$Z^3 = 0$</td>
</tr>
<tr>
<td>$* p = \frac{1}{5}$</td>
<td>$Z^8 - Z^3 = 0$</td>
</tr>
<tr>
<td>$p &gt; \frac{1}{5}$</td>
<td>$Z^8 = 0$</td>
</tr>
</tbody>
</table>

The $p < 0$ limit equation, $-1 = 0$, is absurd, so the $p < 0$ scalings are excluded. The limits for $0 < p < \frac{1}{5}$ and $p > \frac{1}{5}$ give only $Z = 0$, which is their way of saying that they don’t resolve any roots: For instance, if the root $z(\epsilon)$ is bounded as $\epsilon \to 0$, the corresponding $Z(\epsilon) = e^p z(\epsilon)$ converges to zero for any $p > 0$. Hence, scalings with $p > 0$ are too coarse to resolve it. The limit equation $Z^8 = 0$ for $p > \frac{1}{5}$ indicates that scalings with $p > \frac{1}{5}$ are too coarse to resolve any roots at all. This leaves the so-called distinguished limits with $p = 0$ or $p = \frac{1}{5}$. The $p = 0$ limit equation is the same as (1.2), which is obtained by setting $\epsilon = 0$ in (1.1). This nails the three roots which converge to the cube roots of $-1$. The $p = \frac{1}{5}$ scaling nails the 5 diverging roots: The limit equation $Z^8 - Z^3 = 0$ has zero as a triple root. This corresponds to the three bounded roots which appear to vanish in this scaling. The remaining 5 roots are the 5-th roots of 1,

$$Z = e^{\frac{2i\pi k}{5}}, \ k = 0, 1, 2, 3, 4.$$  

The corresponding approximations to the roots $z(\epsilon)$ of (1.1) are

$$z(\epsilon) \simeq \epsilon^{-\frac{1}{5}} e^{\frac{2i\pi k \epsilon}{5}}.$$  

The circles in Figure 1.1 mark the approximations (1.5) for the same sequence of $\epsilon$’s, $\epsilon = 2^{-k}$, $k = 1, \ldots 10$ while the dots are the actual solutions.

Singularly perturbed polynomial equations often appear as the characteristic equations of linear, constant coefficient ODE’s. Initial and boundary value problems involving such ODE’s are singularly perturbed as well. The famous paradoxes of how charged particles interact with their own radiation can be discussed in this context.
Radiation reaction

The power radiated into the electromagnetic field by a particle of charge \( q \), moving along a non-relativistic trajectory \( \mathbf{x} = \mathbf{x}(t) \) is given by a well-known formula due to Lamour,

\[
p = \frac{2}{3} \frac{q^2}{c^3} |\ddot{\mathbf{x}}|^2. \tag{1.6}
\]

Here, \( c \) is the speed of light. The heart of the calculation leading to (1.6) is the retarded solution of the wave equation with a point source moving along \( \mathbf{x} = \mathbf{x}(t) \) and evaluation of the electromagnetic energy flux through a small sphere about \( \mathbf{x} = \mathbf{x}(t) \).

How does the outgoing radiation effect the particle motion? The simplest classical model proposes a “radiation reaction force” \( \mathbf{f}(t) \) so the rate of work \( \mathbf{f}(t) \cdot \dot{\mathbf{x}} \) done by this force on the particle is negative, and balances the dissipation rate (1.6). Specifically, we seek \( \mathbf{f}(t) \) so

\[
\mathbf{f}(t) \cdot \dot{\mathbf{x}}(t) = -p = -\frac{2}{3} \frac{q^2}{c^3} |\ddot{\mathbf{x}}(t)|^2. \tag{1.7}
\]

It can’t really be done: In the right-hand side,

\[
|\ddot{\mathbf{x}}|^2 = -\dddot{\mathbf{x}} \cdot \dddot{\mathbf{x}} + \frac{1}{2} (|\dot{\mathbf{x}}|^2)^{\ldots}
\]

so the proposed balance (1.7) becomes

\[
\mathbf{f} \cdot \dddot{\mathbf{x}} = \frac{2}{3} \frac{q^2}{c^3} \dddot{\mathbf{x}} \cdot \dddot{\mathbf{x}} - \frac{1}{3} \frac{q^2}{c^3} (|\dot{\mathbf{x}}|^2)^{\ldots}. \tag{1.8}
\]

The first term on the right-hand side of (1.8) suggests the identification

\[
\mathbf{f} = \frac{2}{3} \frac{q^2}{c^3} \dddot{\mathbf{x}}, \tag{1.9}
\]

but the second term on the right-hand side remains unaccounted for.

The usual argument is to restrict the discussion to periodic motions \( \mathbf{x}(t) \) and look at the average of the energy balance (1.8) over a period. This conveniently does away with the second term in (1.7). In any case, (1.9) summarizes the standard model in which the particle’s own outgoing radiation is supposed to induce a radiation force upon itself.

As a specific example, consider the radiation damping of a particle of mass \( m \) and charge \( q \) in a uniform magnetic field of strength \( B \). Motion
in a plane perpendicular to the field is possible. Let \( z(t) \) be the complex function of time whose real and imaginary parts are components of velocity. The equation of motion consistent with the given magnetic field and the standard model of radiation force is
\[
\dot{m} z + \frac{qB}{c} i z - \frac{2 q^2}{3 c^3} \ddot{z} = 0. \tag{1.10}
\]
The dimensionless form of (1.10) using \( mc \) as the unit of time is
\[
\dot{z} + i z - \epsilon \ddot{z} = 0, \tag{1.11}
\]
where \( \epsilon \) is the dimensionless parameter
\[
\epsilon := \frac{2}{3} \frac{q^3 B}{m^2 c^4}. \tag{1.12}
\]
For an electron in a lab strength field \( B = 5000 \) gauss, \( \epsilon \) is tiny, \( \epsilon = 10^{-12} \).

For \( \epsilon = 0 \), the solutions of (1.11) are proportional to \( e^{-it} \), corresponding to the well-known circular motion of a particle in a uniform magnetic field and no radiation damping. Now turn on the radiation damping, \( 0 < \epsilon \ll 1 \). There are complex exponential solutions of (1.11)
\[
z = e^{\sigma t},
\]
where \( \sigma \) satisfies the characteristic equation
\[
\epsilon \sigma^2 - \sigma - i = 0. \tag{1.13}
\]
This polynomial equation is singularly perturbed in the sense of the preceding example. The roots, and their expansions in powers of \( \epsilon \) are
\[
\sigma_1 = \frac{1 - \sqrt{1 + 4 \epsilon}}{2 \epsilon} = -i - \epsilon + \ldots, \tag{1.14}
\]
\[
\sigma_2 = \frac{1 + \sqrt{1 + 4 \epsilon}}{2 \epsilon} = \frac{1}{\epsilon} + i \ldots .
\]
As \( \epsilon \to 0 \), \( \sigma_1 \) converges to the root \(-i\) of the reduced \( \epsilon = 0 \) equation, and \( \sigma_2 \) diverges like \( \frac{1}{\epsilon} \). The solution \( z = e^{\sigma_1 t} \) is approximated by
\[
z(t) \simeq e^{-\epsilon t} e^{-i t}. \tag{1.15}
\]
Chapter 1. What is a singular perturbation?

This represents an orbit that slowly spirals to the origin. That is what the theory of radiation force is supposed to describe. But the solution

\[ z(t) = e^{\sigma_2 t} \simeq e^{i \epsilon t} \]  

(1.16)

with violent exponential growth as \( \epsilon \to 0 \) is a disaster. Can one adopt a “don’t ask, don’t tell” policy, and simply exclude the growing solutions from analysis of “physical” problems? As a simple example, suppose there is no magnetic field for \( t < 0 \), during which time the particle is at rest. For times \( t > 0 \), an incident electromagnetic wave propagating in the \( x_1 \) direction induces a uniform magnetic field in the \( x_3 \) direction and a uniform electric field in the \( x_2 \) direction, in a neighborhood of the particle. For \( t > 0 \), the equation of motion is

\[ \dot{z} + i z - i - \epsilon \ddot{z} = 0. \]

(The electric field is represented by the constant term \(-i\).) The solution with \( z(0) \) and \( \dot{z}(0) \) both zero (\( z \) and \( \dot{z} \) are continuous across \( t = 0 \)) is

\[ z(t) = 1 - \frac{\sigma_2 e^{\sigma_1 t} - \sigma_1 e^{\sigma_2 t}}{\sigma_2 - \sigma_1}. \]  

(1.17)

The growing term proportional to \( e^{\sigma_2 t} \) can be approximated by

\[ -i \epsilon e^{\frac{t}{\epsilon}} e^{it}. \]

As \( \epsilon \to 0 \), the coefficient \( i \epsilon \to 0 \), but clearly growth is only delayed. And not by much: The growing term dominates the solution when

\[ \epsilon e^{\frac{t}{\epsilon}} \gg 1 \quad \text{or} \quad t \gg \epsilon \log \frac{1}{\epsilon}. \]

The right-hand side vanishes as \( \epsilon \to 0 \), so disaster is almost instantaneous.

The breakdown of the standard model of radiation reaction has been examined intermittently throughout the 20th century. In 1938, P.A.M. Dirac attempted to re-examine carefully the energy-momentum balance between a charged “point” particle and its retarded field. He was ambushed by the non-integrable energy-momentum singularity along the particle world line. He found that the infinity could be formally cancelled, leaving the radiation force (1.9) as a residual, if the particle interacted with a combination of its advanced and regarded fields. So now the electron according to Dirac has memories of the future! He proposed that this future memory could
be used to suppress violent self acceleration of charged particles in ordinary macroscopic reality. In the preceding example with electric and magnetic fields turned on at $t = 0$, Dirac’s procedure is to allow only the damped solution in $t > 0$, and uphold continuity of $z$ and $\dot{z}$ at $t = 0$ with a small component of growing solution in $t < 0$. This leads to

$$z(t) = \begin{cases} \frac{\epsilon\sigma_1}{1 - \epsilon\sigma_1} e^{\frac{t}{\epsilon}} & t < 0, \\ 1 - \frac{1}{1 - \epsilon\sigma_1} e^{\sigma_1 t} & t \geq 0. \end{cases} \quad (1.18)$$

Notice that this solution asymptotes to uniform motion $z = 1$ as $t \to \infty$. A simple approximation to (1.18) based on the approximation $\sigma_1 \simeq -i - \epsilon$ is

$$z(t) = \begin{cases} i\epsilon e^{\frac{t}{\epsilon}} & t < 0, \\ 1 - (1 - i\epsilon) e^{-\epsilon t} e^{-it} & t \geq 0. \end{cases} \quad (1.19)$$

Figure 1.3 depicts the trajectory of $z(t)$ based on (1.19). The only hint of that notion, “memories of the future”, is the tiny kink in the trajectory near the starting point $z = 0$. 

![Figure 1.3: The trajectory of z(t) based on (1.19)](image-url)
Chapter 1. What is a singular perturbation?

**Convection-diffusion boundary layer**

The function \( c(x, \epsilon) \) defined in \( x \geq 0 \) satisfies the boundary value problem

\[
\epsilon c_x + \phi'(x)c = f \quad \text{in} \quad x > 0, \tag{1.20}
\]

\[
c = 0 \quad \text{at} \quad x = 0, \tag{1.21}
\]

Here, \( \epsilon > 0 \) and \( f > 0 \) are given constants, and \( \phi(x) \) is a given function. The solution \( c(x, \epsilon) \) is examined in the limit \( \epsilon \to 0 \). A physical interpretation is helpful in understanding it: Think of \( c \) as the concentration of particles in a solution. \( \phi(x) \) is the potential energy of a particle, so \( -\phi'(x) \) is the force on it. The force causes a particle to drift with velocity proportional to \( -\phi'(x) \). If this is the only way the particles move, the rate at which they cross into \((0, x)\) would be minus the product of their velocity and density at \( x \), proportional to \( \phi'(x)c(x, \epsilon) \). Transport of particles by a given velocity field is called convection. In addition there is a diffusion of particles induced by random molecular motions at finite temperature. The rate at which particles enter \((0, x)\) due to diffusion is proportional to \( c_x(x, \epsilon) \). The net rate or influx of particles entering \((0, x)\) is the sum of convective and diffusive components. The quantity

\[
\epsilon c_x + \phi'c
\]

which appears in the ODE (1.20) is a non-dimensionalized influx, and \( \epsilon \) is recognized as a dimensionless diffusion coefficient. Under steady conditions, the rate at which particles cross into \((0, x)\) must have a uniform value \( f \) independent of \( x \). This is what the ODE (1.20) says. Otherwise, there would be subintervals of \( x > 0 \) in which particles would accumulate over time, and the density couldn’t be time independent. The zero boundary condition at \( x = 0 \) represents absorption of particles, once they cross \( x = 0 \).

The solution of the boundary value problem (1.20), (1.21) is

\[
c(x, \epsilon) = \frac{f}{\epsilon} \int_0^x \exp \left( \frac{\phi(s) - \phi(x)}{\epsilon} \right) ds. \tag{1.22}
\]

The behavior of \( c(x, \epsilon) \) as \( \epsilon \to 0 \) is inferred from this exact solution.

The most straightforward limit process is \( \epsilon \to 0 \) with \( x > 0 \) fixed. Specifically, define

\[
c^0(x) := \lim_{\epsilon \to 0} c(x, \epsilon). \tag{1.23}
\]

It seems reasonable that \( c^0(x) \) should satisfy the reduced equation

\[
\phi'(x)c^0(x) = f,
\]
obtained by setting $\epsilon = 0$ in (1.22). This would give

$$c^0(x) = \frac{f}{\phi'(x)}.$$  

(1.24)

Assume for now that $\phi'(x)$ is uniformly positive in $x > 0$, so $c^0(x)$ is well defined in $x > 0$. The result (1.24) in fact follows from the integral representation (1.22). The systematic approximation of integrals like (1.22) as $\epsilon \to 0$ will be examined in more detail later. But the essential idea is captured by a formal calculation: Figure 1.4 is a qualitative graph of the exponential in (1.22) as a function of $s$ for $0 < \epsilon \ll 1$, the exponential is much less than 1 over most of the interval $0 < s < x$, and rises steeply to value 1 as $s \to x^-$. Notice that

$$\left. \frac{d}{ds} \left( \exp \left( \frac{\phi(s) - \phi(x)}{\epsilon} \right) \right) \right|_{s=x} = \frac{1}{\epsilon} \phi'(x)$$

so it appears that most of this rise happens in an interval of thickness $\epsilon$. This motivates a change of variable in the integral (1.22). The new variable in place of $s$ is

$$S := \frac{x - s}{\epsilon}.$$ 

The integral (1.22) is rewritten as

$$c = f \int_0^x \exp \left( \frac{\phi(x - \epsilon S) - \phi(x)}{\epsilon} \right) dS,$$
and the formal $\epsilon \to 0$ limit is

$$f \int_0^\infty \exp(-\phi'(x)S)dS = \frac{f}{\phi'(x)} = c^0(x).$$

In general, $c^0(x)$ does not satisfy the zero boundary condition at $x = 0$:

$$c^0(0) = \frac{f}{u} \neq 0, \quad u := \phi'(0). \quad (1.25)$$

This is an indication that the convergence of $c(x, \epsilon)$ to $c^0(x)$ as $\epsilon \to 0$ is non-uniform at $x = 0$. Figure 1.5 depicts a numerical approximation of $c(x, \epsilon)$ for

![Figure 1.5: Numerical solution $c(x, \epsilon)$ (dark line) and inner and outer solutions](image)

specific choices

$$\phi(x) = x + \frac{x^3}{3}, \quad f = 1, \quad \epsilon = .1.$$
A graph of $c^0(x)$ is included for comparison. It is seen that $c(x, \epsilon)$ undergoes a rapid increase from zero in a thin boundary layer to the right of $x=0$.

The width of the boundary layer goes to zero as $\epsilon \to 0$. How does the width scale with $\epsilon$? How does the magnitude of $c$ in the boundary layer scale with $\epsilon$? The mathematical approach to these questions is to introduce a representation of $c(x, \epsilon)$ in the boundary layer that embodies scalings of $c$ and $x$ with respect to $\epsilon$:

$$c(x, \epsilon) = \epsilon^{-q} C(X := \epsilon^{-p} x, \epsilon). \quad (1.26)$$

The exponents $q$ and $p$ are yet to be determined. It is assumed that $C(X, \epsilon)$ converges to a limit function $C^0(X)$ as $\epsilon \to 0$ with $X$ fixed. The ideas embodied in (1.26) are that the boundary layer thickness is proportional to $\epsilon^p$ and the magnitude of $c$ in the boundary layer is proportional to $\epsilon^{-q}$.

Substitution of (1.26) into the ODE (1.21) gives

$$\epsilon^{1-q-p} C_X + \epsilon^{-q} \phi'(\epsilon^p X) C = f. \quad (1.27)$$

For each choice of $q, p$ one gets a reduced equation by taking the limit of (1.27) as $\epsilon \to 0$. The process is analogous to the analysis of the polynomial equation (1.1). In particular, there are distinguished limits characterized by balances between two or more terms of (1.27) in powers of $\epsilon$. For instance the reduced equation $\phi'(x)c^0(x) = f$ that characterizes the solution in the limit $\epsilon \to 0$ with $x > 0$ fixed is obtained with $q = 0, p = 0$. There is another distinguished limit that balances all three terms of (1.27). If $u := \phi'(0) > 0$, the limit equation with $q = 0, p = 1$ is

$$C^0_X + u C^0 = f. \quad (1.28)$$

It is plausible that $q = 0, p = 1$ is the boundary layer limit. $p = 1$ means that the boundary layer thickness is proportional to $\epsilon$. One might have guessed $q = 0$ at the outset, since the numerical solution in Figure 1.5 suggests that the net rise of $c(x, \epsilon)$ in the boundary layer appears to converge to the positive value $c^0(0) = \frac{\xi}{u}$ as $\epsilon \to 0$.

The conclusion that boundary layer thickness is proportional to $\epsilon$ can be argued by traditional dimensional analysis. Think of (1.20) as an equation in dimensional quantities. In particular, $\epsilon$ is a diffusion coefficient with units of $(\text{length})^2 \div \text{time}$, and the drift velocity at $x = 0$, $u = \phi'(0)$ has units of length $\div \text{time}$. The unique length formed from $\epsilon$ and $u$ is $\frac{\epsilon}{u}$. Notice that $c$
has units of \( 1 \div \text{length} \), but its magnitude does not enter into the balance of convection and diffusion, which are both linear in \( c \).

Having established, by fair means or foul, that boundary layer thickness is proportional to \( \epsilon \), the solution (1.22) in the boundary layer can be examined rigorously: Introduce the scaled displacement,

\[
X := \frac{x}{\epsilon},
\]

and define

\[
C(X, \epsilon) := c(\epsilon X, \epsilon) = \int_0^X \exp \left( \frac{\phi(\epsilon X - \epsilon S) - \phi(\epsilon X)}{\epsilon} \right) dS. \tag{1.29}
\]

\( C(X, \epsilon) \) is the exact solution (1.22) expressed as a function of \( X \) and \( \epsilon \). The integral representation (1.29) is obtained from (1.22) by substituting \( x = \epsilon X \), and changing the integration variable from \( s \) to \( S := X - \frac{s}{\epsilon} \). Since \( x \) has magnitude \( \epsilon \) in the boundary layer, the appropriate limit process for the boundary layer is \( \epsilon \to 0 \) with \( X := \frac{x}{\epsilon} \) fixed. Under this limit process, the exponent in (1.29) has limit

\[
\lim_{\epsilon \to 0} \frac{\phi(\epsilon X - \epsilon S) - \phi(\epsilon X)}{\epsilon} = -uS,
\]

and

\[
C^0(X) := \lim_{\epsilon \to 0} C(X, \epsilon) = f \int_0^X e^{-uS} dS = \frac{f}{u} (1 - e^{-uX}). \tag{1.30}
\]

\( C^0(X) \) is displayed as a function of \( x \) in Figure 1.5. Notice that \( C^0(X) \) satisfies the boundary condition \( C^0(0) = 0 \) but generally does not converge to \( c^0(x) \) in the limit \( \epsilon \to 0 \), \( x > 0 \) fixed.

In summary, two complimentary limits of the exact solution \( c(x, \epsilon) \) have been found. One is \( c^0(x) \), in the limit \( \epsilon \to 0 \) and \( x > 0 \) fixed. The other is \( C^0(X) \), in the limit \( \epsilon \to 0 \), \( X := \frac{x}{\epsilon} \) is fixed. The limits describe the solution in the complimentary intervals, outside and inside the boundary layer. Traditionally, \( c^0(x) \) is called an outer approximation, and \( C^0(X) \) an inner approximation. How can the outer and inner approximations be “joined” to form an approximation to \( c(x, \epsilon) \) uniformly valid in \( x > 0 \)? First, observe a certain compatibility between \( c^0(x) \) and \( C^0(X) \) called matching. From (1.25) and (1.30) it follows that

\[
\lim_{X \to \infty} C^0(X) = \frac{f}{u} \lim_{x \to 0} c^0(x). \tag{1.31}
\]
Chapter 1. What is a singular perturbation?

A traditional verbal characterization of the matching is: “Outer limit of inner approximation = Inner limit of outer approximation”. The common value $\frac{f}{u}$ of the limits in (1.31) is called the “common part of the inner and outer approximations”. The matching (1.31) informs the sought for “joining” of inner and outer approximations. Define

$$c^u(x, \epsilon) := c^0(x) + C^0(X) - \frac{f}{u} = \frac{f}{u} \left\{ \frac{u}{\phi'(x)} - e^{-\frac{u}{\epsilon}} \right\}. \quad (1.32)$$

This sum of inner and outer approximations minus the common part is the candidate for a uniform approximation to $c(x, \epsilon)$ in $0 < x < 1$. Figure 1.6 compares $c^u(x, \epsilon)$ to the numerical solution.

Figure 1.6: Numerical solution (dark line) compared with $c^u$
Chapter 1. What is a singular perturbation?

Modulated oscillations

The function $x = x(t, \epsilon)$ defined in $t \geq 0$ satisfies the initial value problem

$$\ddot{x} + \frac{x}{(1 + \epsilon t)^2} = 0, \quad t > 0 \tag{1.33}$$

$$x(0) = 0, \quad \dot{x}(0) = 1, \tag{1.34}$$

where $\epsilon > 0$ is a given positive constant. Physically, (1.33) represents a harmonic oscillator whose natural frequency $\omega$ depends on time,

$$\omega = \frac{1}{1 + \epsilon t}. \tag{1.35}$$

The exact solution is

$$x(t, \epsilon) = a \sin \xi, \tag{1.36}$$

where the amplitude $a$ and phase $\xi$ are given by

$$a = a(t, \epsilon) = \sqrt{\frac{1 + \epsilon t}{1 - \frac{\epsilon^2}{4}}}, \tag{1.37}$$

$$\xi = \xi(t, \epsilon) = \sqrt{1 - \frac{\epsilon^2 \log(1 + \epsilon t)}{4\epsilon}}. \tag{1.38}$$

Figure 1.7a is the graph of $x(t, \epsilon)$ for $\epsilon = .01$ in $0 < t < 300$. The convergence of $x(t, \epsilon)$ as $\epsilon \to 0$ is non-uniform in $0 < t < \infty$: The limit of (1.36) as $\epsilon \to 0$ with $t > 0$ fixed is $x^0(t) = \sin t$. The graph of the residual $x(t, \epsilon) - x^0(t)$ in Figure 1.7b displays oscillations that grow with increasing $t$. In fact, let $\{t_n\}$ be the sequence of maxima of $\sin \xi(x, t)$, so

$$\xi(t_n, \epsilon) = \sqrt{1 - \frac{\epsilon^2 \log(1 + \epsilon t_n)}{4\epsilon}} = \frac{\pi}{2} + 2\pi n$$

for $n = 0, 1, 2, \ldots$. By elementary estimology, $|x(t_n, \epsilon) - x^0(t)| > e^{n\pi\epsilon} - 1 \to \infty$ as $n \to \infty$. In summary, the initial value problem (1.33), (1.34) is singularly perturbed.

Heuristically, it is clear what goes on: In the limit $\epsilon \to 0$, $x(t, \epsilon)$ converges uniformly to a sinusoidal function of $t$ over any subinterval of $0 < t < \infty$ that contains a fixed number of oscillations, independent of $\epsilon$. But after a large number of oscillations, on the order of $\frac{1}{\epsilon}$, the amplitude and frequency
Chapter 1. What is a singular perturbation?

Figure 1.7a

Figure 1.7b
of the oscillations undergo large relative changes. This kind of non-uniform convergence is called "modulated oscillation".

We carry out an $\epsilon \to 0$ limit process on the solution that displays the "local periodicity" over a few cycles, and how the periodic structure changes after many cycles, on the order of $\frac{1}{\epsilon}$. In the expressions (1.37), (1.38) for amplitude and phase, time appears scaled by $\epsilon$, so it’s natural to examine a limit with the “slow time” $\epsilon t$ near a fixed value $T$. To focus on a few cycles of oscillation about $t = \frac{T}{\epsilon}$, we introduce time with origin shifted to $\frac{T}{\epsilon}$. That is,

$$\tau := t - \frac{T}{\epsilon}. \quad (1.39)$$

We express the amplitude and phase (1.37), (1.38) in terms of $\tau$,

$$a = \sqrt{\frac{1 + T + \epsilon \tau}{1 - \frac{\epsilon^2}{4}}},$$

$$\xi = \sqrt{1 - \frac{\epsilon^2 \log(1 + T + \epsilon \tau)}{4 \epsilon}}.$$  

In the limit $\epsilon \to 0$ with $\tau$ fixed,

$$a \to \sqrt{1 + T}, \quad \xi - \frac{\log(1 + T)}{\epsilon} \to \frac{\tau}{1 + T}. \quad (1.40)$$

The limit amplitude is independent of the shifted time $\tau$, so in an interval of time about $\frac{T}{\epsilon}$ with duration independent of $\epsilon$, the amplitude of oscillations is uniformly approximated by a constant (which happens to be $\sqrt{1 + T}$). So we say, “the local amplitude at slow time $T$ is $\sqrt{1 + T}$”. If we go to another interval with a different slow time $T$, we’ll get a different amplitude. In the second of equations (1.40), $\frac{\log(1 + T)}{\epsilon}$ approximates the phase at time $\frac{T}{\epsilon}$. What’s really important is the phase difference $\xi - \frac{\log(1 + T)}{\epsilon}$: In the limit $\epsilon \to 0$, $\tau$ fixed, we see that its time rate of change is the constant $\frac{1}{1+T}$. So we say: “The local frequency at slow time $T$ is $\frac{1}{1+T}$”. That’s what a physicist expects by simple inspection of the original ODE (1.33). Again: different $T$’s, different frequencies.

Problems whose solutions exhibit modulated oscillations are sometimes tractable by constructive approximation methods. Like the “inner and outer approximations” that we discovered in the boundary layer example, they acknowledge the two or more characteristic times that live in these problems.
Chapter 1. What is a singular perturbation?

But unlike boundary layer problems, the two or more characteristic times are “everywhere intermingled”, so the constructive methods have names like “averaging” or “multiple scales”.

Here we present a constructive “physicist approximation” to the amplitude $a$ in (1.36). Let’s generalize the ODE (1.33) to

$$\ddot{x} + \Omega^2(\epsilon t) x = 0,$$  \hspace{1cm} (1.41)

where the natural frequency is a given function $\Omega(T)$ of the slow time $T := \epsilon t$. For a few cycles of oscillation about the “long time” $\frac{T}{\epsilon}$, we expect a nearly uniform sinusoidal oscillation. In terms of the shifted time $\tau$ in (1.39), we’d say

$$x \rightarrow a(T) \cos \Omega(T) \tau$$  \hspace{1cm} (1.42)

as $\epsilon \rightarrow 0$ with $\tau$ fixed. We dropped an additive constant from the phase: it can be absorbed by shifting the origin of time from $\frac{T}{\epsilon}$. We simply “put in” the expected local frequency $\Omega(T)$, which is a given function of slow time $T$. The analysis we’re about to do determines the slow time dependence of amplitude, $a = a(T)$.

The slow change of natural frequency over long time does cumulative work on the oscillator, and there is slow cumulative change in the energy

$$e := \frac{1}{2} \dot{x}^2 + \frac{\Omega^2}{2} x^2.$$  \hspace{1cm} (1.43)

We compute

$$\dot{e} = \epsilon \Omega \frac{d\Omega}{dT} x^2.$$  \hspace{1cm} (1.44)

Terms $x\ddot{x}$ and $\Omega^2 x\dot{x}$ cancel by ODE (1.41), and indeed we see that change in energy is induced specifically by slow time dependence of $\Omega(T := \epsilon t)$. The time interval $\frac{T}{\epsilon} < t < \frac{T}{\epsilon} + \frac{2\pi}{\Omega(T)}$ corresponds to almost one cycle of oscillation, so the change of energy in this cycle is approximately

$$\delta e \simeq \epsilon \Omega \frac{d\Omega}{dT} \int_{\frac{T}{\epsilon}}^{\frac{T}{\epsilon} + \frac{2\pi}{\Omega}} x^2 dt.$$  \hspace{1cm} (1.45)

The factor $\Omega \frac{d\Omega}{dT}$ occurs outside the integral, since its nearly constant in $\frac{T}{\epsilon} < t < \frac{T}{\epsilon} + \frac{2\pi}{\Omega}$. The integral itself is approximated using the local approximation (1.42) to $x$, and changing the variable of integration to shifted time $\tau$. Again,
Chapter 1. What is a singular perturbation?

We can express the right-hand side in terms of $e$: Substituting the local approximation (1.42) of $x$ into (1.43) (and again treating $a(T)$ and $\Omega(T)$ as approximate constants), we find that over a single period near time $t = T/\epsilon$, the energy is an approximate constant,

$$e \simeq \frac{1}{2} \Omega^2 a^2. \quad \text{(1.47)}$$

Setting $a^2 \simeq \frac{2}{\Omega^2} \epsilon e$ in (1.46), we see that

$$\delta e \simeq 2\pi \epsilon \frac{1}{\Omega^2} \frac{d\Omega}{dT} e. \quad \text{(1.48)}$$

So far, (1.48) tracks the discrete change in energy over a single cycle near time $T/\epsilon$. Now the last and rather subtle idea: We propose that the leading approximation to the energy is a nice function of slow time $e(T)$, and $\delta e$ in (1.48) is the differential of this energy due to the differential $dT = \epsilon \frac{2\pi}{\Omega}$ of slow time corresponding to a single period. That is

$$\delta e \simeq 2\pi \epsilon \frac{1}{\Omega} \frac{de}{dT}. \quad \text{(1.49)}$$

Comparison of (1.48), (1.49) gives

$$\frac{1}{\epsilon} \frac{de}{dT} \simeq \frac{1}{\Omega} \frac{d\Omega}{dT} \quad \text{(1.50)}$$

and we see that $e(T)$ is proportional to $\Omega(T)$. We recover $a(T)$ from (1.47) with $e$ proportional to $\Omega$, giving $a(T)$ proportional to $\frac{1}{\sqrt{\Omega(T)}}$. For our original example (1.33) with $\Omega = \frac{1}{1+T}$, this gives $a$ proportional to $\sqrt{1+T}$, as we’ve already seen from the exact solution.

This “physicist analysis” has a superficial appearance of being “elementary”, but on closer examination, there are subtle insights and guesses. The force must be with you.
Chapter 1. What is a singular perturbation?
Chapter 2

Asymptotic expansions

Let \( f(\epsilon) \) be a function of \( \epsilon \). A common-sense notion of “an approximation \( f^0(\epsilon) \) to \( f(\epsilon) \) for small \( \epsilon \)” is that

\[
f(\epsilon) - f^0(\epsilon) = f^0(\epsilon)(1 + e(\epsilon))
\]

(2.1)

where the “relative error” \( e(\epsilon) \) vanishes in the limit \( \epsilon \to 0 \). This rudimentary sense of approximation is implicit in the examples of the preceding section. A notion of successive approximations, in which the relative error is greatly reduced at each stage, has not yet been formulated.

A conventional notion of successive approximations is embodied in convergent power series. Assume \( f(\epsilon) \) is analytic in an interval \(|\epsilon| < R\), and is represented there by the power series

\[
f(\epsilon) = a_0 + a_1 \epsilon + a_2 \epsilon^2 \ldots .
\]

(2.2)

(2.2) means that the partial sums

\[
f_n(\epsilon) := a_0 + a_1 \epsilon + \cdots + a_n \epsilon^n
\]

(2.3)

converge to \( f(\epsilon) \) as \( n \to \infty \) for each fixed \( \epsilon \) in \(|\epsilon| < R\). A complimentary notion of approximation is to examine the error in approximating \( f(\epsilon) \) by a partial sum with fixed \( n \) as \( \epsilon \to 0 \). With this in mind, rewrite (2.2) in the form

\[
f(\epsilon) - f_n(\epsilon) = \epsilon^{n+1} r_n(\epsilon)
\]

(2.4)

where \( r_n(\epsilon) \) is the analytic function in \(|\epsilon| < R\) represented by convergent series

\[
r_n(\epsilon) \equiv a_{n+1} + a_{n+2} \epsilon + a_{n+3} \epsilon^2 \ldots .
\]

(2.5)
Let \( k \) be the largest integer less than or equal to \( n \) with \( a_k \neq 0 \). Then for \(|\epsilon|\) sufficiently small, \( a_k \epsilon^k \) is the smallest non-zero term in \( f_n(\epsilon) \). This smallest term is bigger than the error in the sense that

\[
\frac{f(\epsilon) - f_n(\epsilon)}{a_k \epsilon^k} = \epsilon^{n-k+1} \frac{r_n(\epsilon)}{a_k} \to 0
\]
as \( \epsilon \to 0 \). Hence, each partial sum “approximates \( f(\epsilon) \) for small \( \epsilon \)” because its smallest term becomes much bigger than the error as \( \epsilon \to 0 \). This is the essential notion of an asymptotic expansion.

The notion of an asymptotic expansion of a function \( f(\epsilon) \) begins with a sequence of sums in the form,

\[
f_n(\epsilon) := \sum_{1}^{n} a_k \delta_k(\epsilon)
\] (2.6)

for \( n = 1 \ldots N \). Here, the \( a_k \) are given constants and \( \delta_k(\epsilon) \) are given “gauge functions” that “decrease with \( k \)” in the sense that

\[
\delta_{k+1}(\epsilon) = e_k(\epsilon) \delta_k(\epsilon)
\] (2.7)

where \( e_k(\epsilon) \) are functions which vanish as \( \epsilon \to 0 \). There is a convenient shorthand to indicate the order relation (2.7) without explicit mention of the functions \( e_k(\epsilon) \). Specifically,

\[
f(\epsilon) = o(g(\epsilon))
\] (2.8)

means that there is \( e(\epsilon) \) which vanishes as \( \epsilon \to 0 \) so that

\[
f(\epsilon) = e(\epsilon)g(\epsilon).
\] (2.9)

If (2.9) holds, but \( e(\epsilon) \) is only bounded as \( \epsilon \to 0 \), the notation is

\[
f(\epsilon) = O(g(\epsilon)).
\] (2.10)

A sequence of gauge functions \( \{\delta_k(\epsilon)\} \) with

\[
\delta_{k+1}(\epsilon) = o(\delta_k(\epsilon))
\]
is called an asymptotic sequence. Power series are based on the asymptotic sequence \( 1, \epsilon, \epsilon^2 \ldots \). In practice, other asymptotic sequences arise naturally. For instance, an example at the end of this section treats a solution
Chapter 2. Asymptotic expansions

of Laplace’s equation due to a varying source along a line in $\mathbb{R}^3$ as the distance $r$ from the line goes to zero. There arises naturally the asymptotic sequence $1, \log r, r^2 \log r, r^4 \log r, \ldots$, due to the logarithmic character of the fundamental solution in two dimensions.

The sum (2.6) with $n = N$ is called an $N$-term asymptotic expansion of the function $f(\epsilon)$ if for each $n$, $1 \leq n \leq N$,

$$f_n(\epsilon) - f(\epsilon) = o(\delta_n(\epsilon)).$$

(2.11)

As anticipated before, the basic idea is that the error is smaller than the smallest term retained in the sum. If there are constants $a_k$ defined for all $k \geq 1$, so that the corresponding sums (2.6) satisfy the order relation (2.11) for all $n \geq 1$, the infinite sequence of sums $f_1(\epsilon), f_2(\epsilon), \ldots$ is called an asymptotic expansion of $f(\epsilon)$, and the notation is

$$f(\epsilon) \sim \sum_{n=1}^{\infty} a_n \delta_n(\epsilon) \quad \text{as} \quad \epsilon \to 0.$$  

(2.12)

Why is the notion of an asymptotic series, as opposed to convergent series, so important? In real applications of perturbation theory, the calculation of all coefficients $a_1, a_2, \ldots$ in (2.6) is often out of the question. Getting the first few terms, or even just the leading term might be an enormous effort. But due to the nature of (2.6) as an asymptotic series, even a one-term or two-term expansion has a small relative error if $\epsilon$ is sufficiently small. In addition, there are divergent asymptotic series. In these cases, adding more terms with $\epsilon$ fixed is useless.

A divergent but asymptotic series

Define

$$f(\epsilon) := \int_0^{\infty} \frac{e^{-x}}{1+x} dx.$$  

(2.13)

Now substitute into (2.13) the algebraic identity

$$\frac{1}{1+x} = 1 - x + x^2 + \cdots + (-x)^n + \frac{(-x)^{n+1}}{1+x}$$

and carry out term-by-term integrations to find

$$f(\epsilon) = f_n(\epsilon) + e_n(\epsilon),$$  

(2.14)
where
\[ f_n(\epsilon) = \epsilon - \epsilon^2 + 2\epsilon^3 - \cdots + (-1)^n n! \epsilon^{n+1}. \] (2.15)

and
\[ e_n(\epsilon) = (-1)^{n+1} \int_0^{\infty} \frac{x^{n+1}}{1 + x} e^{-\frac{x}{\epsilon}} \, dx. \] (2.16)

It readily follows from (2.16) that
\[ |e_n(\epsilon)| < \int_0^{\infty} x^{n+1} e^{-\frac{x}{\epsilon}} \, dx = (n + 1)! \epsilon^{n+2}. \] (2.17)

It now follows from (2.14) that
\[ |f_n(\epsilon) - f(\epsilon)| < (n + 1)! \epsilon^{n+2}, \] (2.18)

so
\[ f_n(\epsilon) - f(\epsilon) = o(\epsilon^{n+1}) \]

for all \( n \geq 0 \). Hence, \( f(\epsilon) \) has the asymptotic expansion
\[ f(\epsilon) \sim \sum_{0}^{\infty} (-1)^n n! \epsilon^{n+1} \] (2.19)

which is clearly divergent for all \( \epsilon > 0 \).

Given a divergent asymptotic expansion, it is natural to ask what number of terms gives the smallest error for fixed “small” \( \epsilon \)? In the preceding example, there is the simple upper bound (2.17) on the error in the \( n + 1 \)-term expansion. It is easy to estimate the value of \( n \) which minimizes this upper bound. For small \( \epsilon \), this value of \( n \) is presumably large, and by use of Stirling’s approximation,
\[ n! \sim n^n e^{-n} \sqrt{2\pi n} \] (2.20)
as \( n \to \infty \), it follows that
\[ \log \{(n + 1)! \epsilon^{n+2}\} \sim n \log n - n - n \log \frac{1}{\epsilon}. \]

The right-hand side is minimized for \( n = \frac{1}{\epsilon} \), and the resulting approximation to the minimum of \( (n + 1)! \epsilon^{n+2} \) is \( e^{-\frac{1}{\epsilon}} \). Figure 2.1 is a graph of \((n + 1)! \epsilon^{n+2}\) vs. \( n \) for \( \epsilon = .1 \).
Asymptotic expansions — the usual suspects

The analysis of the preceding example is a foretaste of Laplace’s method, which generates asymptotic expansions of integrals with strong exponentials. The starting point is integrals of the type

$$f(\epsilon) = \int_a^b g(x)e^{-h(x)/\epsilon} \, dx. \quad (2.21)$$

Without loss of generality we can take $a = 0, b = 1$.

First, take $h(x) = x$, and $g(x)$ continuous in $[0, 1]$, with a convergent power series in $|x| < R$. No restriction on $R$ except that it is positive. Then the asymptotic power series of $f(\epsilon)$ in (2.21) is

$$f(\epsilon) \sim g(0)\epsilon + g'(0)\epsilon^2 + \ldots g^{(n)}(0)\epsilon^{n+1} + \ldots. \quad (2.22)$$

The formal calculation leading to (2.22) is substitution of the power series for $g(x)$ into (2.21) and term-by-term integration over the extended interval

Figure 2.1: A graph of $(n + 1)!\epsilon^{n+2}$ vs. $n$ for $\epsilon = .1$
[0, ∞) It might seem peculiar that we can “get away with this”, especially when \( R < 1 \), but we can as the actual proof now shows:

Choose \( \rho, 0 < \rho < \min(1, R) \), and write (2.21) as

\[
\int_0^\rho g(x)e^{-\frac{x}{\epsilon}}dx + \int_\rho^1 g(x)e^{-\frac{x}{\epsilon}}dx.
\]  

(2.23)

In the integral over \([0, \rho]\), we substitute for \( g(x) \) a partial sum of its power series plus remainder

\[
g(x) = g_n(x) + r_n(x)
\]  

(2.24)

where

\[
g_n(x) := g(0) + g'(0)x + \cdots + \frac{g^{(n)}(0)}{n!}x^n
\]  

(2.25)

and

\[
r_n(x) := \frac{g^{(n+1)}(\zeta(x))}{(n+1)!}x^{n+1}.
\]  

(2.26)

In (2.26), \( \zeta(x) \) is a function with \( 0 \leq \zeta(x) \leq x \). The result of the indicated substitution can be arranged to read

\[
f(\epsilon) = \int_0^\epsilon g_n(x)e^{-\frac{x}{\epsilon}}dx 
+ \int_0^\rho r_n(x)e^{-\frac{x}{\epsilon}}dx
+ \int_\rho^1 g(x)e^{-\frac{x}{\epsilon}}dx 
- \int_\rho^\infty g_n(x)e^{-\frac{x}{\epsilon}}dx.
\]  

(2.27)

Notice that the first integral on the right-hand side is over the whole interval \( x \geq 0 \) and this is compensated by the appearance of \( g_n(x) \) in the last integral over \( x \geq \rho \). The first integral on the right-hand side can be done explicitly and is in fact the partial sum of the asymptotic series (2.22). Hence, (2.27) becomes

\[
f(\epsilon) = g(0)\epsilon - g'(0)\epsilon^2 - \cdots g^{(n)}(0)\epsilon^{n+1} = \int_0^\epsilon r_n(x)e^{-\frac{x}{\epsilon}}dx 
+ \int_\rho^1 g(x)e^{-\frac{x}{\epsilon}}dx
- \int_\rho^\infty g_n(x)e^{-\frac{x}{\epsilon}}dx.
\]  

(2.28)
To show that $f(\epsilon)$ is asymptotic to the series (2.22), it remains to show that the right-hand side of (2.28) is $O(\epsilon^{n+1})$ as $\epsilon \to 0$. First, $r_n(x)$ in (2.26) is bounded in absolute value by $R_n x^{n+1} (n+1)!$, where $R_n$ is an upper bound on $|g^{(n+1)}(\zeta(x))|$ in $0 \leq x \leq \rho$. Then

$$\left| \int_0^\rho r_n(x)e^{-\frac{x}{\epsilon}}dx \right| \leq \frac{R_n}{(n+1)!} \int_0^\infty x^{n+1}e^{-\frac{x}{\epsilon}}dx$$

$$= R_n \epsilon^{n+2} = O(\epsilon^{n+2}) = o(\epsilon^{n+1}).$$

The second integral on the right-hand side of (2.28) is bounded in absolute value by

$$M \epsilon e^{-\frac{\rho}{\epsilon}} = o(\epsilon^{n+1})$$

where $M$ is an upper bound on $|g(x)|$ in $[0,1]$. The last integral on the right-hand side of (2.28) is a linear combination of components $\int_\rho^\infty x^k e^{-\frac{x}{\epsilon}}dx$, with $k = 0, 1, \ldots n$. By elementary substitutions, we have

$$\int_\rho^\infty x^k e^{-\frac{x}{\epsilon}}dx = \rho^{k+1} e^{-\frac{\rho}{\epsilon}} \int_0^\infty e^{-\frac{\rho y + ky}{\epsilon}}dy.$$

For $0 < \rho < 1$, $\epsilon < \frac{\rho}{k}$, the right-hand side is less than or equal to

$$e^{-\frac{\rho}{\epsilon}} \int_0^\infty e^{-\frac{\rho y + ky}{\epsilon}}dy = \frac{e^{-\frac{\rho}{\epsilon}}}{\frac{\rho}{\epsilon} - k} = o(\epsilon^{n+1}).$$

In summary, the whole right-hand side of (2.28) is $o(\epsilon^{n+1})$.

The next increment of generalization is to take $h(x)$ to be a smooth, increasing function in $[0,1]$, with $h'(x)$ bounded below by a positive constant. This case reduces to the previous: Introduce

$$y = H(x) := h(x) - h(0) \quad (2.29)$$

as the variable of integration in place of $x$, and (2.21) (with $a = 0$, $b = 1$) expressed as a $y$-integral is

$$f(\epsilon) = e^{-\frac{h(0)}{\epsilon}} \int_0^{H(1)} G(y) e^{-\frac{y}{\epsilon}}dy \quad (2.30)$$

where

$$G(y) = \frac{g(H^{-1}(y))}{H'(H^{-1}(y))}. \quad (2.31)$$
Assuming $G(y)$ has a power series about $y = 0$ with positive radius of convergence, we quickly obtain

$$f(\epsilon) \sim e^{-\frac{h(0)}{\epsilon}} \{ G(0) \epsilon + G'(0) \epsilon^2 + G''(0) \epsilon^2 + \ldots \}.$$  \hspace{1cm} (2.32)

Let’s look at the first term: Compute from (2.31),

$$G(0) = \frac{g(0)}{h'(0)}$$

so the one-term expansion of $f(\epsilon)$ is

$$f(\epsilon) \sim \frac{\epsilon g(0)}{h'(0)} e^{-\frac{h(0)}{\epsilon}}.$$  \hspace{1cm} (2.33)

This approximation blows up as $h'(0) \to 0$, signaling a “regime change”. $h'(x)$ uniformly positive is the “generic case” of $h(x)$ with global minimum at $x = 0$. $h'(x)$ uniformly negative is the “generic case” of global minimum at $x = 1$, and clearly that’s nothing new. The natural next step is to assume that the global minimum $x = x_m$ of $h(x)$ is in the open interval $(0, 1)$, with $h''(x_m) > 0$. In this case, we modify the change of variable in (2.29). Specifically, assume that

$$h(x) - h(x_m) = (x - x_m)^2 k(x)$$

with $k(x)$ positive and analytic in some interval $|x - x_m| \leq \rho$ about $x = x_m$. Notice $k(x_m) = \frac{h''(x_m)}{2} > 0$. The proposed change of variable is

$$y = H(x) := (x - x_m) \sqrt{k(x)}$$  \hspace{1cm} (2.34)

for $|x - x_m| \leq \rho$. The “worst case” is when $\rho$ is so small that this interval is completely inside $[0, 1]$. This is the one we’ll do: We convert the $x$-integral over $[x_m - \rho, x_m + \rho]$ into a $y$-integral using the change of variable (2.34), and for now, leave the $x$-integral over the compliment $I := [0, 1] - (x_m - \rho, x_m + \rho)$ alone. We obtain

$$f(\epsilon) = e^{-\frac{h_m}{\epsilon}} \int_{H(x_m - \rho)}^{H(x_m + \rho)} G(y)e^{-\frac{y^2}{2}} dy + \int_{I} g(x)e^{-\frac{h(x)}{\epsilon}} dx.$$  \hspace{1cm} (2.35)
Chapter 2. Asymptotic expansions

Here, \( h_m := h(x_m) \), and \( G(y) \) is given by (2.31) with the new \( H(x) \) in (2.34). In the \( y \)-integral, we insert a power series for \( G(y) \), extend the \( y \) interval to \((-\infty, \infty)\) and we do term-by-term integration to obtain

\[
e^{-\frac{h_m}{\epsilon}} \sqrt{\pi \epsilon} \left\{ G(0) + \frac{G''(0)}{2^2 1!} \epsilon + \frac{G^{(4)}(0)}{2^4 2!} \epsilon^2 + \ldots \right\}.
\]

(2.36)

The second integral on the right-hand side of (2.35) is bounded in absolute value by \( Me^{-\frac{h_I}{\epsilon}} \), where \( M \) is an upper bound on \(|g(x)|\) in \([0, 1]\) and \( h_I \) is the minimum value of \( h(x) \) in \( I \). If \( h_m < h_I \), then each term of (2.36) is larger than \( Me^{-\frac{h_I}{\epsilon}} \), and (2.36) is in fact the asymptotic expansion of \( f(\epsilon) \). Other accessory details (truncation of power series, extension of \( y \) integral to \((-\infty, \infty)\)) are going to be very similar to the original analysis of the first case leading to (2.22).

The most useful part of (2.36) is the leading term. Given \( G(y) \) in (2.31) and \( H(x) \) as in (2.34), we compute

\[
G(0) = \sqrt{\frac{2}{h''_m}} g_m
\]

where \( h''_m =: h''(x_m) \), \( g_m =: g(x_m) \), and the one-term expansion of \( f(\epsilon) \) is

\[
f(\epsilon) \sim \sqrt{\frac{2\pi \epsilon}{h''_m}} e^{-\frac{h_m}{\epsilon}} g_m.
\]

(2.37)

In summary, the asymptotic expansion of the integral (2.21) is controlled by the global minimum of \( h(x) \) in \([a, b]\). The expressions (2.32), (2.36) are representative of the two “generic” cases: Minimum at an endpoint (say \( x = a \) with \( f'(a) > 0 \)), or minimum at \( x_m \) in \((a, b)\) with \( f''(x_m) > 0 \).

**Steepest descent method**

Physical problems motivate the generalization from real integrals (2.21) to complex contour integrals

\[
f(\epsilon) = \int_C g(z)e^{\frac{h(z)}{\epsilon}} dz,
\]

(2.38)

where \( g(z) \) and \( h(z) \) are complex analytic functions. For instance, linear PDE which are translation invariant in space and time variables \( x \) and \( t \)
have “plane wave” solutions $e^{ikx+\sigma(k)t}$, where $k$ is real and $\sigma(k)$ is a complex-valued function of $k$, called the PDE’s dispersion relation. More general solutions are given by superpositions of the form

$$u(x,t) = \int_{-\infty}^{\infty} \hat{u}(k) e^{ikx+\sigma(k)t} dk.$$  \hfill (2.39)

We recognize $\hat{u}(k)$ as the Fourier transform of $u(x,0)$. To examine the long-time limit $t \to \infty$ with $\frac{x}{t}$ fixed, we write

$$u(x,t) = \int_{-\infty}^{\infty} \hat{u}(k) e^{t(\sigma(k)+\frac{ix}{t}k)} dk,$$  \hfill (2.40)

which looks like a special case of (2.38) with $\epsilon = \frac{1}{t}$ and the contour $C$ is the real axis. If $\hat{u}(k)$ and $\sigma(k)$ are analytic, the contour can be suitably deformed to expedite the proposed asymptotic evaluation of (2.39).

For instance, the Fresnel integral

$$f(\epsilon) := \int_{-\infty}^{\infty} e^{\frac{i\epsilon z^2}{\epsilon}} dz$$  \hfill (2.41)

is evaluated by deforming the contour from the real axis to the diagonal line with parametric representation $z = e^{i\pi t}$, $t$ real, so

$$f(\epsilon) = \int_{-\infty}^{\infty} e^{-\frac{i\epsilon t^2}{\epsilon}} (e^{i\pi t} dt) = \sqrt{\frac{\pi}{\epsilon}} e^{i\frac{\pi}{4}} = \sqrt{\frac{\pi}{2\epsilon}} (1+i).$$

The diagonal line has a clear relation to the real part of the exponent, $\text{Re} \left( \frac{iz^2}{\epsilon} \right) = -\frac{2t^2}{\epsilon}$. Figure 2.2 is its contour map in the $z$ plane. $z = 0$ is a saddle point, and the diagonal line is the “steepest descent” path from it, everywhere parallel to the gradient of $\text{Re} \left( \frac{iz^2}{\epsilon} \right)$. 
Chapter 2. Asymptotic expansions

Figure 2.2

The notion of “steepest descent” path is key to the asymptotic evaluation of (2.38) as $\epsilon \to 0$. Since $h(z) = u(z) + iv(z)$ is a complex analytic function, the “generic” critical point $z_*$ of $\text{Re } h(z) = u(z)$ is a saddle. The “steepest descent” path is a curve which passes through $z_*$, and is everywhere parallel to $u_x + i u_y$. If $z(t)$ is a parametric representation of the steepest descent path, we have $u_y \dot{x} - u_x \dot{y} = 0$. By the Cauchy–Riemann equations, $v_x = -u_y$, $v_y = u_x$, so $v_x \dot{x} + v_y \dot{y} = 0$, and it follows that $v(z) = v(z_*)$, or $\text{Im } h(z) = \text{Im } h(z_*) = \text{constant}$ along the steepest descent path. If $C$ in (2.38) is in fact the steepest descent path, we have $h(z) - h(z_*)$ real along $C$, with maximum value zero at $z = z_*$. Now write (2.38) as

$$e^{h(z_*)} \int_{-\infty}^{\infty} g(z(t)) z'(t) e^{\frac{h(z(t)) - h(z_*)}{\epsilon}} dt, \quad (2.42)$$

where $h_* := h(z_*)$. We recognize $e^{h(z(t)) - h_*}$ as a real exponential that strongly decreases to zero away from $t = t_*$ with $z(t) = z_*$. Hence the real and imaginary parts of the complex contour integral both fit in the framework of the real integral (2.21).

So here is the drill: Find the saddle points of $\text{Re } h(z)$ by solving $h'(z) = 0$. Next, deform the original contour $C$ into the steepest descent path through a saddle point $z_*$. In many cases, Cauchy’s theorem guarantees that the original path can be deformed to the steepest descent path, with no change
in the value of the integral. But be careful: If in deforming the path you
cross an isolated singularity, you might “pick up a residue”. Once the original
contour integral is reduced to an integral over a steepest descent path, we
are in the framework of the real integral (2.21).

Here is a typical example of the steepest descent method: \( u(x,t) \) is the
fundamental solution of the PDE

\[
  u_t = \frac{1}{3} u_{xxx}
\]

in \(-\infty < x < \infty, \ t > 0\) which satisfies \( u(x,t) \to \delta(x) \) as \( t \to 0 \). The Fourier
integral representation of \( u(x,t) \) is

\[
  u(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(kx - \frac{1}{3} t^3)} dk.
\]  

(2.44)

We determine the long-time asymptotic behavior, \( t \to \infty \) with \( v := \frac{\pi}{t} \) fixed. Hence write (2.44) as

\[
  u(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(kv - \frac{1}{3} z^3)} dk = \frac{1}{\pi} \text{Re} \int_{0}^{\infty} e^{ith(z)} dz
\]

(2.45)

where

\[
  h(z) := i \left( vz - \frac{z^3}{3} \right).
\]

(2.46)

Saddle points are zeros of \( h'(z) = i(v - z^2) \) so \( z = \pm \sqrt{v} \). If \( v > 0 \), they are
on the real axis, if \( v < 0 \), on the imaginary. Hence the analysis naturally
divides into \( v > 0 \), and \( v < 0 \).

\( v > 0 \)

The steepest descent path through \( z = \sqrt{v} \) is determined from \( \text{Im} h(z) = \text{Im} h(\sqrt{v}) \). Write it out explicitly,

\[
  vz - \frac{x^3}{3} + xy^2 = \frac{2}{3} v^{\frac{3}{2}}
\]

or

\[
  y^2 = \frac{1}{3x}(x + 2\sqrt{v})(x - \sqrt{v})^2
\]

(2.47)

in \( x > 0 \). The branch of (2.47) corresponding to descent from \( z = \sqrt{v} \)
(\( \text{Re} h(z) \) decreasing as we move away from \( \sqrt{v} \)) is

\[
  y = -\sqrt{\frac{x + 2\sqrt{v}}{3x}} (x - \sqrt{v}).
\]

(2.48)
In Figure 2.3, this curve oriented by decreasing $y$ is labeled $C$. Here is a simple way to see why the steepest descent path has $y$ decreasing in $x$: Near $z = \sqrt{v}$,

$$h(z) - h(\sqrt{v}) \sim \frac{h''(\sqrt{v})}{2}(z - \sqrt{v})^2 = -i\sqrt{v}(z - \sqrt{v})^2$$

and the right-hand side is real and negative along the line with $z - \sqrt{v}$ proportional to $e^{-\frac{1}{4\pi}}$. We can also explain why $C$ has the asymptotic line $y = -\frac{x}{\sqrt{3}}$: In the limit $|z| \to \infty$, $h(z) \sim -\frac{iz^3}{\sqrt{3}}$ and $\text{Re} h(z) \sim \frac{r_3}{3} \sin 3\theta$ ($r, \theta$ polar coordinates of $z$). The contour plot of $\frac{r^3}{3} \sin 3\theta$ appears in Figure 2.4. Since the original path in the integral (2.45) is the positive real axis, starting from $z = 0$, we will also need the steepest descent path from $z = 0$. That’s easy: It is the positive imaginary axis $\left(h(iy) = -(xy + \frac{y^3}{3})\right)$. In Figure 2.3, we label the positive imaginary axis with upward orientation by $C'$. Evidently, we are going to replace the original path (positive real axis) by $C' + C$.

The integral along $C'$ is

$$\frac{1}{\pi} \text{Re} \int_0^\infty e^{t h(iy)} d(iy) = \frac{1}{\pi} \text{Re} \int_0^\infty e^{-t(vy + \frac{x^3}{3})} d(iy) = 0.$$
This leaves the integral along \( C \), which we write as

\[
\frac{1}{\pi} \text{Re} \left\{ e^{t h(\sqrt{v})} \int_C e^{t(h(z)-h(\sqrt{v}))} dz \right\}.
\] (2.49)

Here, \( h(\sqrt{v}) = \frac{2}{3}i v^{\frac{3}{2}} \), and along \( C \),

\[
h(z) - h(\sqrt{v}) = -y \left( v + \frac{y^2}{3} - x^2 \right),
\]

which is real. Hence, (2.49) becomes

\[
\frac{1}{\pi} \text{Re} \left\{ e^{\frac{2}{3}i v^{\frac{3}{2}}} \int_C e^{-ty(\psi + \frac{y^2}{3} - x^2)} dz \right\}.
\] (2.50)

Once we parametrize \( C \), we end up with two real integrals, to which we apply Laplace’s method. For instance (2.48) defines \( x \) as an implicit function of \( y \), and we could write

\[
\int_C e^{-ty(\psi + \frac{y^2}{3} - x^2)} dz = \int_{-\infty}^{\infty} e^{-ty(\psi + \frac{y^2}{3} - x^2)} (x'(y) + i) dy.
\] (2.51)

The leading approximation to (2.51) is

\[
\int_{-\infty}^{\infty} e^{-2\sqrt{vt}y^2} (-1 + i) dy = \sqrt{\frac{\pi}{2\sqrt{vt}}} (1 - i) = \sqrt{\frac{\pi}{\sqrt{vt}}} \sqrt{1 - \frac{i}{4}}.
\]
and the corresponding (leading) approximation to \( u(x,t) \) as \( t \to \infty \) with \( v = \frac{\pi}{t} \) fixed and positive is

\[
    u(x,t) \sim \sqrt{\frac{1}{\pi \sqrt{vt}}} \cos \left( \frac{2}{3}tv^{\frac{3}{2}} - \frac{\pi}{4} \right).
\]

(2.52)

Here, we used \( x(y) - \sqrt{v} \sim -y \) as \( y \to 0 \).

\( v < 0 \)

The saddle points are at \( z_s = \pm i \sqrt{|v|} \), and the corresponding values of \( h(z_s) \) are \( \pm \frac{2}{3}|v|^{\frac{3}{2}} \). You might think that the “saddle at \( z_s = i \sqrt{|v|} \) wins”, and the asymptotic approximation to \( u \) is proportional to the large exponential \( e^{\frac{2}{3}|v|^{\frac{3}{2}}t} \). But that’s not the way of the Force. First, look at the steepest descent path from \( z = 0 \): Since \( h(0) = 0 \), the steepest descent path has

\[
    0 = \text{Im} h(z) = x \left( y^2 - \frac{x^2}{3} - |v| \right).
\]

(2.53)

Since we start from \( z = 0 \), we must go along the imaginary axis \( (x = 0) \). On \( x = 0 \), \( h(iy) = |v|y - \frac{x^3}{3} \), so the steepest descent path from \( z = 0 \) “falls” to the “wrong” saddle point \( z_s = -i \sqrt{|v|} \). This path from \( z = 0 \) to \( z = -i \sqrt{|v|} \) is labeled \( C' \) in Figure 2.5. Now its clear what happens next:
Chapter 2. Asymptotic expansions

We have $\text{Im} \ h(-i \sqrt{|v|}) = 0$, and from (2.53) we see that $\text{Im} \ h(z)$ is also zero along the hyperbola $y^2 - \frac{x^2}{3} = |v|$. We simply follow the portion of hyperbola (labeled $C$ in Figure 2.5) that goes down “into the valley of monkey saddle” (Figure 2.4) whose “bottom” is the line $y = -\frac{x}{\sqrt{3}}$. In summary, the deformed path is $C' + C$. As in the $v > 0$ case, the contribution from the $C'$ integral is exactly zero. The $t \to \infty$ limit of the $C$-integral contains the exponentially small prefactor $e^{-h(z_*)} = e^{-\frac{2}{3} |v|^{\frac{3}{2}} t}$. The bottom line: “Not much happening in $x > 0$”.

A primer on linear waves

Let’s summarize: In the limit $t \to \infty$, $v := \frac{x}{t}$ fixed, $u(x,t)$ is exponentially small for $v < 0$, and the leading approximation for $v > 0$ is

$$u \sim \sqrt{\frac{1}{\pi \sqrt{vt}}} \cos \left( \frac{2}{3} v^{\frac{3}{2}} t - \frac{\pi}{4} \right). \tag{2.54}$$

The limit process can be reformulated as $\epsilon \to 0$ with $T := \epsilon t > 0$ and $X := \epsilon x$ fixed. In (2.54) we introduce $X, T$ in place of $x, t$ to obtain

$$u \sim \sqrt{\epsilon} a(X, T) \cos \left( \frac{\theta(X, T)}{\epsilon} - \frac{\pi}{4} \right) \tag{2.55}$$
as $\epsilon \to 0$ with $T > 0$, $X$ fixed. Here, the amplitude $a(X, T)$ and scaled phase $\theta(X, T)$ are given by

$$a(X, T) = \frac{1}{\sqrt{\pi}} (XT)^{-\frac{1}{4}},$$

$$\theta(X, T) = \frac{2}{3} X^{\frac{3}{2}} T^{-\frac{1}{2}}. \tag{2.56}$$

Clearly, (2.55) exhibits modulated oscillations in space and time in the sense described in Chapter 1.

Let’s look at the local structure of a few space or time oscillations about $(x, t) = \left( \frac{X_0}{\epsilon}, \frac{T_0}{\epsilon} \right)$ for some fixed values of $X_0, T_0$. Introduce translated space and time variables,

$$\sigma := x - \frac{X_0}{\epsilon}, \quad \tau := t - \frac{T_0}{\epsilon}.$$

In the limit $\epsilon \to 0$ with $\sigma, \tau$ fixed, $a \to a(X_0, T_0)$, and

$$\theta - \frac{2}{3} X_0^{\frac{3}{2}} T_0^{-\frac{1}{2}} \to \theta_X(X_0, T_0) \sigma + \theta_T(X_0, T) \tau.$$
In this limit, the wavefield (2.55) looks like a uniform plane wave, with uniform amplitude $a(X_0, T_0)$, wavenumber $\theta_X(X_0, T_0)$ and frequency $-\theta_T(X_0, T_0)$. In this sense, $a(X, T)$ is called the *local amplitude*, and

$$k(X, T) := \theta_X(X, T),$$
$$\omega(X, T) := -\theta_T(X, T)$$

(2.57)

are called the *local wavenumber* and *local frequency*. For the specific phase in (2.56),

$$k = \left(\frac{X}{T}\right)^\frac{1}{2}, \quad \omega = \frac{1}{3} \left(\frac{X}{T}\right)^3 = \frac{1}{3} k^3.$$  

(2.58)

An observer who starts out from $(X, T) = (0, 0)$ with uniform velocity $v = \frac{X}{T} > 0$ sees only one wavenumber $k = \sqrt{v}$, and one frequency $\omega = \frac{1}{3} k^3$, which is what you get from the dispersion relation for plane wave solutions of the original PDE (2.43).

We can understand these results from a larger perspective. Let $\theta(X, T)$ be the phase of a wave whose local wavenumber and frequency (2.57) satisfy the dispersion relation

$$\omega = W(k).$$

(2.59)

From the definitions (2.57) it follows that the wavenumber is locally conserved,

$$k_T + \omega_X = \theta_{XT} - \theta_{TX} = 0.$$ 

Now put $\omega = W(k)$ and we get a Burger’s-like equation

$$k_T + (W(k))_X = 0,$$ 

(2.60a)

or

$$k_T + W'(k)k_X = 0.$$ 

(2.60b)

Notice that the “total number of waves” between $X = X_1$ and $X = X_2$ is

$$\frac{1}{2\pi} \int_{X_1}^{X_2} k(\epsilon x, T) dx = \frac{1}{2\pi \epsilon} \int_{X_1}^{X_2} k(X, T) dX,$$

and by (2.60), has time rate of change

$$- \left[ \frac{1}{2\pi \epsilon} W(k) \right]_{X_1}^{X_2}.$$
Chapter 2. Asymptotic expansions

so \( \frac{1}{2\epsilon} W(k) \) is the “flux of waves”.

Let \( X = X(T) \) be a level curve of a solution to the PDE (2.60) for \( k(X,T) \). We have

\[
0 = \frac{d}{dT} k(X(T),T) = (k_T + \dot{X} k_X) = (\dot{X} - W(k)) k_X,
\]

so level curves have (constant) velocities

\[
\dot{X} = W'(k).
\]

Physicists call \( W'(k) \) the group velocity and the lines in \( X, T \) plane of constant \( k \) are called group lines. If the wavefield originates near \((x,t) = (0,0)\), the group lines originate from \((X,T) = (0,0)\) so they are given by

\[
\frac{X}{T} = W'(k). \tag{2.61}
\]

For \( W(k) = \frac{1}{3} k^3 \), (2.61) reduces to the first of equations (2.58).

The amplitude \( a(X,T) \) satisfies a local conservation law. We can adopt (2.55) as an anzatz for solutions of the PDE \( u_t = \frac{1}{3} u_{xxx} \), or even more simply, let’s set

\[
u = a(X,T) e^{i\theta(X,T)} . \tag{2.62}
\]

We get real solutions by taking real or imaginary parts of (2.62). Due to linearity we ignore the prefactor \( \sqrt{\epsilon} \) in (2.55), and we also blew off the phase shift \(-\frac{\pi}{4}\). We calculate

\[
u_t = (-i\omega a + \epsilon a_T) e^{i\theta}, \quad \nu_{xxx} = \{-ik^3 a - 3\epsilon(a_X k^2 + a k k_x) + O(\epsilon^2)\} e^{i\theta}.
\]

Here, we put \( \theta_T = -\omega, \theta_X = k \). Substituting these into the PDE (2.43), we find

\[
w = \frac{1}{3} k^3 + O(\epsilon), \quad a_T + a_X k^2 + a k k_X = O(\epsilon).
\]

In the limit \( \epsilon \to 0 \), the first of these equations reduces to the (expected) dispersion relation, and the second is recast as a local conservation law for \( a^2 \)

\[(a^2)_T + (k^2 a^2)_X = 0.\]
Chapter 2. Asymptotic expansions

We see that $a^2$ is a density convected with the group velocity $k^2$. Notice that there is no convection of $a^2$ in the direction of decreasing $x$ because the group velocity is non-negative. This is consistent with the exponential smallness of $u$ as $t \to \infty$ with $v = \frac{z}{t} < 0$ fixed, as follows from the steepest descent analysis of the Fourier integral representation.

For a general dispersion relation, we expect

$$(a^2)_T + (W'(k)a^2)_X = 0. \quad (2.63)$$

Let’s look at the value $A(T) = a^2(X(T),T)$ along a group line $X = X(T)$ with

$$\dot{X}(T) = W'(k).$$

With the help of (2.63) we calculate

$$\frac{dA}{dT} = (a^2)_T + W'(k)(a^2)_X = -(W'(k))_X A, \quad (2.64)$$

where $(W'(k))_X$ is evaluated at $(X(T),T)$. In the special case of group lines all radiating from $(X,T) = (0,0)$, we have $W'(k) = \frac{X}{T}$ as in (2.61), so (2.64) reduces to $\frac{dA}{dT} = -\frac{A}{T}$ and it follows that $AT = a^2T$ is a time independent constant along a group line. This explains the $\frac{1}{\sqrt{t}}$ prefactor in the result (2.54) from steepest descent asymptotics. The actual spatial dependence of $a^2(X,T)$ at fixed $T$ is determined by the initial wavefield at $t = 0$. This is explained in Problem 2.9.

A hard logarithmic expansion

The electrostatic potential due to a line change along the $z$-axis in $\mathbb{R}^3$ is given by

$$\phi(r,z) = \int_{-\infty}^{\infty} \frac{\sigma(z')}{{(z-z')^2 + r^2}^{\frac{1}{2}}} dz'. \quad (2.65)$$

Here, $r$ is distance from the $z$-axis. $\sigma(z')$ is line charge density at $z = z'$ on $z$ axis. $\sigma(z')$ is assumed to be analytic and integrable. The factor

$$\frac{1}{{(z-z')^2 + r^2}^{\frac{1}{2}}}$$

in (2.65) is the electrostatic potential of a unit charge at $(0,0,z')$ seen at any point with elevation $z$ relative to the plane $z = 0$ and distance $r$ from the $z$-axis. Hence, (2.65) is the standard superposition of point charge fields.
Chapter 2. Asymptotic expansions

The asymptotic expansion of the potential $\phi$ is $r \to 0$ is examined here. The form of the expansion can be discerned from the Laplace equation,

$$\Delta \phi = \phi_{rr} + \frac{1}{r} \phi_r + \phi_{zz} = 0 \quad (2.66)$$

satisfied by $\phi$. If the source $\sigma(z)$ is sinusoidal, $\sigma(z) = e^{ikz}$, then $\phi$ takes separation of variables from

$$\phi = F(r)e^{ikz}$$

and $F(r)$ satisfies the modified Bessel equation

$$F'' + \frac{1}{r}F' - k^2 F = 0.$$ 

There is a regular singular point at $r = 0$, and the standard local analysis about $r = 0$ establishes that $F(r)$ takes the form

$$a(r) \log r + b(r), \quad (2.67)$$

where $a(r)$ and $b(r)$ are analytic in $r^2$. The general source $\sigma(z)$ is a superposition of sinusoidal sources, and the local structure (2.67) generalizes to

$$\phi(r, z) = a(r, z) \log r + b(r, z) \quad (2.68)$$

where $a$ and $b$ are analytic in $r^2$ and $z$. It is now clear that the asymptotic sequence for the $r \to 0$ expansion is going to be

$$\log r, \ 1, \ r^2 \log r, \ r^2, \ r^4 \log r, \ r^4, \ldots \ . \quad (2.69)$$

Let us examine in more detail the information about $a(r, z)$ and $b(r, z)$ that follows from the Laplace equation (2.66). Substituting representation (2.68) of $\phi$ into Laplace’s equation (2.66) gives a pair of PDE’s for $a(r, z)$ and $b(r, z)$. The craftsmanship of the details shall not be slighted here. First, compute $\Delta \phi$ in terms of $a$ and $b$:

$$\nabla \phi = a \hat{r} + \log r \nabla a + \nabla b$$

and

$$\Delta \phi = \nabla \cdot \nabla \phi = 2 \nabla a \cdot \frac{\hat{r}}{r} + a \nabla \cdot \left( \frac{\hat{r}}{r} \right) + \log r \Delta a + \Delta b.$$
In the above, \( \hat{r} \) is the radial unit vector, \( \nabla a \cdot \hat{r} = a_r \) and \( \nabla \cdot (\hat{r} \bigtriangleup) = 0 \) so
\[
\bigtriangleup \phi = \log r \bigtriangleup a + \frac{2a_r}{r} + \bigtriangleup b = 0. \tag{2.70}
\]
Since \( a \) and \( b \) are analytic in \( r^2 \), \( \frac{2a_r}{r} + \bigtriangleup b \) has no \( \log r \) component and cancellation of \( \log r \) component from (2.70) requires that
\[
\bigtriangleup a = a_{rr} + \frac{1}{r} a_r + a_{zz} = 0. \tag{2.71}
\]
It now follows that
\[
\bigtriangleup b = b_{rr} + \frac{1}{r} b_r + b_{zz} = -\frac{2a_r}{r}. \tag{2.72}
\]
Now substitute into (2.71) a series for \( a(r, z) \) in powers of \( r^2 \),
\[
a \sim a^0(z) + a^2(z) r^2 + \ldots. \tag{2.73}
\]
All the \( a^{2k}(z) \), \( k > 0 \) are determined from \( a^0(z) \) and the expansion of \( a(r, z) \) as \( r \rightarrow 0 \) takes the form
\[
a \sim \left\{ 1 - \frac{r^2}{(2 \cdot 1)!} (\partial_z)^2 + \frac{r^4}{(2^2 \cdot 2)!} (\partial_z)^4 + \ldots \right\} a^0(z). \tag{2.74}
\]
Substituting into (2.72) the asymptotic series (2.73) for \( a \) and a power series for \( b \) similar to (2.73), and solving the resulting recursion equations yields for \( b \),
\[
b \sim \left\{ 1 - \frac{r^2}{(2 \cdot 1)!} (\partial_z)^2 + \frac{r^4}{(2^2 \cdot 2)!} (\partial_z)^4 + \ldots \right\} b^0(z) \\
+ \left\{ \frac{r^2}{(2 \cdot 1)!} (\partial_z)^2 - \frac{1}{2} \frac{r^4}{(2^2 \cdot 2)!} (\partial_z)^4 + \frac{1}{3} \frac{r^6}{(2^3 \cdot 3)!} (\partial_z)^6 + \ldots \right\} a^0(z). \tag{2.75}
\]
In summary, the asymptotic series for \( \phi(r, z) \) is (2.68) with \( a(r, z) \) and \( b(r, z) \) determined from \( a^0(z) \) and \( b^0(z) \) according to (2.74), (2.75).

The determination of \( a^0(z) \) and \( b^0(z) \) comes from hard analysis of the integral representation (2.65) of \( \phi \). Some preliminary changes of integration variable area great enablers. First, (2.65) can be rewritten as
\[
\phi(r, z) = \int_0^\infty \frac{\sigma(z + z') + \sigma(z - z')}{(z'^2 + r^2)^{\frac{1}{2}}} dz'. \tag{2.76}
\]
Next, introduce new integration variable $\zeta$ so
\[
d\zeta = \frac{dz'}{(z'^2 + r^2)^{1/2}}.
\] (2.77)

An integral of (2.77) is
\[
z' = r \sinh \zeta.
\] (2.78)

(2.76) becomes
\[
\phi(r, z) = \int_0^\infty \left\{ \sigma(z + r \sinh \zeta) + \sigma(z - r \sinh \zeta) \right\} d\zeta.
\] (2.79)

In the limit $r \to 0$, one must have $\zeta \to \infty$ to maintain fixed values of $r \sinh \zeta$. But $\sinh \zeta \sim \frac{1}{2} e^\zeta$ for $\zeta \gg 1$, and this motivates the last change of variable
\[
v := \frac{r}{2} e^\zeta.
\]

Then
\[
r \sinh \zeta = v - \frac{r^2}{4v}, \quad d\zeta = \frac{dv}{v}
\]
and (2.79) becomes
\[
\phi(r, z) = \int_{\frac{r}{2}}^\infty \left\{ \sigma \left( z + v - \frac{r^2}{4v} \right) + \sigma \left( z - v + \frac{r^2}{4v} \right) \right\} \frac{dv}{v}.
\] (2.80)

Integration by parts gives
\[
\phi(r, z) = -2\sigma(z) \log\frac{r}{2}
- \int_{\frac{r}{2}}^\infty \left\{ \sigma' \left( z + v - \frac{r^2}{4v} \right) - \sigma' \left( z - v + \frac{r^2}{4v} \right) \right\} \left( 1 + \frac{r^2}{4v} \right) \log v \, dv.
\] (2.81)

The integral in (2.81) converges to
\[
\int_0^\infty \{\sigma'(z + v) - \sigma'(z - v)\} \log v \, dv
\]
as $r \to 0$, hence $\phi(r, z)$ has the two-term expansion
\[
\phi(r, z) \sim -2\sigma(z) \log\frac{r}{2} - \int_0^\infty \{\sigma'(z + v) - \sigma'(z - v)\} \log v \, dv
\]
and comparison with

\[ \phi(r, z) \sim a^0(z) \log r + b^0(z) \]

leads to identifications

\[ a^0(z) = -2\sigma(z), \quad (2.82) \]

\[ b^0(z) = (2 \log 2)\sigma(z) - \int_0^\infty \{\sigma'(z + v) - \sigma'(z - v)\} \log v \, dv. \quad (2.83) \]
Chapter 2. Asymptotic expansions
Chapter 3

Matched asymptotic expansions

The notion of matched asymptotic expansions arises from classical problems of continuum mechanics. The most famous is Prandtl’s boundary layer theory. Figure 3.1 shows the streamlines of fluid flow around a “fish” of length $L$ swimming at uniform velocity $-U$ in the $x$ direction, as seen in the fish rest frame. The simplest model of a fluid, called ideal fluid ("dry water", according to Feynman) neglects dissipative stresses and heat transfer. The velocity field within the ideal fluid approximation is tangential, but generally non-zero on the fishes’ skin. Due to this boundary condition, the ideal fluid velocity field inherits $L$ as a characteristic length. In actual fluid ("wet water", according to Feynman), fluid molecules “stick” to the fishes’ skin and the velocity field vanishes there. In a boundary layer adjacent to the fishes’ skin, the formerly neglected functional stresses are strong. The dimensional parameter which quantifies frictional stress is the kinematic viscosity $\nu$, with
units (length)$^2$/time. The thickness of the boundary layer is on the order of

$$\delta = \sqrt{\frac{\nu}{LU}} L$$

(3.1)

We explain the physical ideas behind (3.1). Think of $\nu$ as a “diffusion coefficient of velocity”. If at time zero we have uniform flow over a (fixed) flat plate, the thickness of the boundary layer at time $t > 0$ is on the order of $\sqrt{\nu t}$, as follows from dimensional analysis.\(^1\) This is illustrated in Figure 3.2. In our flow in Figure 3.1, a fluid particle moves characteristic distance $L$ in time $\frac{L}{U}$ so the boundary layer thickness at distance $O(L)$ downstream from the “nose” is $\sqrt{\nu t}$ with $t = \frac{L}{U}$, and that’s $\delta$ in (3.1). The dimensionless ratio $\text{Re} := \frac{LU}{\nu}$ is called the Reynolds’ number. For sufficiently “big and fast fish” (the usual case) $\text{Re} \gg 1$ so $\epsilon := \frac{\delta}{L} = \text{Re}^{-\frac{1}{2}}$ is small.

In dimensionless variables with $L$ as the unit of length and $U$ as unit of velocity the fluid mechanical boundary value problem describing the “fish flow” of Figure 3.1 is singularly perturbed as $\epsilon \to 0$. In the limit $\epsilon \to 0$, $x$ fixed, the velocity field $u$ has an outer asymptotic expansion whose leading term is the “dry water” flow. This expansion is non-uniformly valid as we approach the fishes’ skin, and we turn to the complimentary inner asymptotic expansion of the flow, valid as $\epsilon \to 0$ and $z := \text{normal displacement from skin} = O(\epsilon)$.

The outer and inner expansions exhibit a certain compatibility called matching. Although the inner expansion is non-uniformly valid as $z \to 0$ and

\(^1\)You might object: “The initial condition is ridiculous. You can’t set that up!” But we can quickly accelerate a flat plate from rest in a fluid at rest, and the flow in the “plate frame” looks like Figure 3.2.
Chapter 3. Matched asymptotic expansions

the inner expansion is non-uniformly valid as $\xi \to \infty$, we expect that there is an “intermediate” range of $\xi$, $\epsilon \ll \xi \ll 1$, where neither expansion has “crapped out” entirely, and they actually “agree” with each other. The first job of this chapter is to quantify the “agreement” in precise mathematics.

The vehicle of this exposition is the simple convection-diffusion boundary value problem,

$$\epsilon c_x + \varphi'(x)c = f \quad \text{in} \quad x > 0, \quad c = 0 \quad \text{at} \quad x = 0 \quad (3.2)$$

presented in the introductory chapter. First, recall some preliminary results: We showed that $c(x, \epsilon)$ has an outer limit $c(x, \epsilon) \to c^0(x) := \frac{f}{\varphi'(x)}$ as $\epsilon \to 0$ with $x > 0$ fixed (assuming $\varphi'$ uniformly positive) and an inner limit $c(x, \epsilon) \to C^0(X := \frac{x}{\epsilon}) = \frac{f}{\varphi'(0)}(1 - e^{-X})$ as $\epsilon \to 0$ with $X := \frac{x}{\epsilon} > 0$ fixed. We observed that there is leading order matching expressed by

$$\lim_{x \to 0} c^0(x) = \lim_{X \to \infty} C^0(X) = \frac{f}{\varphi'(0)}.$$

How does “matching” work at higher orders of approximation? Let

$$c(x, \epsilon) = c^0(x) + \epsilon c^1(x) + \epsilon^2 c^2(x) + \ldots \quad (3.3)$$

be the outer asymptotic expansion valid as $\epsilon \to 0$ with $x > 0$ fixed, and

$$c(x, \epsilon) \sim C^0(X) + \epsilon C^1(X) + \epsilon^2 C^2(X) + \ldots \quad (3.4)$$

be the inner expansion, valid as $\epsilon \to 0$ with $X := \frac{x}{\epsilon}$ fixed. We want to explain in what sense the entire inner and outer expansions (3.3), (3.4) “match”. We’ll take $\varphi'(x) = e^x$ and $f = 1$ to simplify the explicit construction of inner and outer expansions so we can examine matching with minimal technical obstruction.

The outer expansion was worked out in Problem 1.2, and we found

$$c^0(x) = e^{-x}, \quad c^1(x) = 1!e^{-2x}, \quad c^2(x) = 2!e^{-3x}, \ldots \quad (3.5)$$

The inner expansion admits a simple term-by-term construction. The ODE in (3.2) in terms of $c = C(X := \frac{x}{\epsilon}, \epsilon)$ is

$$C_x + e^{\epsilon x}C = 1. \quad (3.6)$$
We substitute into (3.6) the asymptotic power series (3.4) to obtain perturbation equations
\[
\begin{align*}
C_X^0 + C^0 &= 1, \\
C_X^1 + C^1 &= -XC^0, \\
C_X^2 + C^2 &= -XC^1 - \frac{X^2}{2} C^0, \ldots.
\end{align*}
\]

The required solutions for \( C^0, C^1, C^2 \) which vanish at \( X = 0 \) are constructed by elementary methods:
\[
\begin{align*}
C^0 &= 1 - e^{-X}, \\
C^1 &= 1 - X - e^{-X} \left( 1 - \frac{X^2}{2} \right), \\
C^2 &= 2 - 2X + \frac{X^2}{2} - e^{-X} \left( 2 - \frac{X^2}{2} - \frac{X^3}{6} + \frac{X^4}{8} \right).
\end{align*}
\]

This exercise can be continued, like Gary Larson’s “Areobics in Hell”.

We see that “leading order matching” as expressed by \( \lim_{x \to 0} c^0(x) = \lim_{X \to \infty} C^0(X) = 1 \) clearly works. But naive generalizations to higher order come to grief immediately: For instance, a proposed matching between two-term outer and inner expansions according to
\[
\lim_{x \to 0} (c^0(x) + \epsilon c^1(x)) = \lim_{X \to \infty} (C^0(X) + \epsilon C^1(X))
\]
is a non-starter because the limit on the right-hand side does not exist. We need to look deeper, at the key idea of overlap domains.

For instance the overlap domain of the leading order outer and inner expansions is the range of \( x \) so their difference is \( o(1) \),
\[
\begin{equation}
\epsilon^0(x) - C^0 \left( X := \frac{x}{\epsilon} \right) = e^{-x} - 1 + e^{-\frac{x}{\epsilon}} = o(1).
\end{equation}
\]
We have \( e^{-\frac{x}{\epsilon}} \ll 1 \) if \( x \gg \epsilon \) and \( e^{-x} - 1 = o(1) \) if \( x = o(1) \). Hence the overlap domain of leading order outer and inner expansions is characterized by
\[
\epsilon \ll x \ll 1.
\]
(3.10) does not represent a definite interval, but rather a range of allowed orders of magnitude of \( x \). We can make it precise like this: \( \epsilon^0(x) - C^0(x) = o(1) \) for \( x = O(\epsilon^p) \) with \( 0 < p < 1 \). Some terminology: We have been
calling the limit processes \( \epsilon \to 0 \) with \( x = O(1) \) and \( X = O(1) \) (equivalently \( x = O(\epsilon) \)) outer and inner limits, so it seems natural to call limit processes \( \epsilon \to 0 \) with \( x = O(\epsilon^p) \), \( 0 < \epsilon < 1 \) intermediate limits.

Now finally we can get a grip on “higher order matching”. For instance, the two-term outer and inner expansions are “matched” if there are intermediate limits (characterized by exponents \( p \) in a subinterval of \((0, 1)\)) so their difference is \( o(\epsilon) \), smaller than the smallest term in either expansion. Here is the drill for determining the overlap domain: First, we write the two-term inner expansion in terms of \( x \) for later comparison with the outer expansion. We drop terms which contain the exponential \( e^{-X} = e^{-\frac{1}{p}x} \) as a factor, since they are exponentially small in any intermediate limit with \( x = O(\epsilon^p) \), \( p < 1 \).

Hence,

\[
C^0(X) + \epsilon C^1(X) = 1 + \epsilon(1 - X) = 1 + \epsilon - x \tag{3.11}
\]

modulo exponentially small terms. Next, we asymptotically expand the two-term outer solution in \( x \) as \( x \to 0 \). We need to do enough terms to reproduce the right-hand side of (3.11), and the residual is indicated only by its order of magnitude:

\[
c^0(x) + \epsilon c^1(x) = 1 - x + O(x^2) + \epsilon + O(\epsilon x) = 1 - x + \epsilon + O(x^2 + \epsilon x). \tag{3.12}
\]

From (3.11), (3.12) we see that the difference between two-term inner and outer expansions is \( O(x^2 + \epsilon x) \). In intermediate limits \( x = O(\epsilon^p) \), this difference is \( o(\epsilon) \) if \( \epsilon^{2p} + \epsilon^{1+p} = o(\epsilon) \) and hence \( p > \frac{1}{2} \). In summary, the overlap domain of two-term outer and inner expansions is \( x = O(\epsilon^p) \), with \( \frac{1}{2} < p < 1 \).

The overlap domain of three-term outer and inner expansions is analyzed similarly: The three-term inner expansion written in terms of \( x \) is

\[
1 + \epsilon(1 - X) + \epsilon^2 \left( 2 - 2X + \frac{1}{2} X^2 \right) = 1 - x + \frac{x^2}{2} + \epsilon(1 - 2x) + 2\epsilon^2, \tag{3.13}
\]

again modulo exponentially small terms. We carry out an \( x \to 0 \) asymptotic expansion of the three-term outer solution. Again, we do enough terms to recover all those in (3.13) and represent residuals by their orders of magnitude. The \( x \to 0 \) behavior of the three-term outer expansion is

\[
1 - x + \frac{x^2}{2} + O(x^3) + \epsilon(1 - 2x + O(x^2)) + \epsilon^2(2 + O(x))
\]

\[
= 1 - x + \frac{x^2}{2} + \epsilon(1 - 2x) + 2\epsilon^2 + O(x^3 + \epsilon x^2 + \epsilon^2 x). \tag{3.14}
\]
From (3.13), (3.14) we see that the difference between three-term inner and outer expansions is $O(x^3 + \epsilon x^2 + \epsilon^2 x)$, and this is $o(\epsilon^2)$ (smaller than the smallest term in either expansion) in intermediate limits with $\epsilon^{3p} + \epsilon^{1+2p} + \epsilon^{2+p} = o(\epsilon^2)$, so $p > \frac{2}{3}$. The upper bound $p < 1$ is still in place to eliminate the exponential terms in the inner expansion. Hence, the overlap domain between three-term outer and inner expansions is $x = O(\epsilon^p)$ where $\frac{2}{3} < p < 1$.

The pattern is now clear: The overlap domain of $n$-term outer and inner expansions is $x = O(\epsilon^p)$ where $\frac{n-1}{n} < p < 1$: If $p < 1$, the terms of inner expansion with factor $e^{-x} = e^{-\frac{x}{\epsilon}}$ are exponentially small. The surviving polynomial component of inner expansion differs from the outer solution by $O(x^n) = O(\epsilon^{np})$ and the residual is $o(\epsilon^{n-1})$ (smaller than the smallest term of either expansion) if $p > \frac{n-1}{n}$.

As the order of matching increases (increasing number of terms in outer and inner expansions) the overlap domain shrinks. This is general. But the number of terms in inner and outer expansions which match are not always equal. This can happen for a variety of reasons. For instance, the asymptotic sequences of gauge functions for the outer and inner expansions may be different.

As a simple example, assume $\varphi'(x)$ in the convection-diffusion problem 3.2 is positive in $x > 0$ but $\varphi'(0) = 0$ with $\varphi''(0) > 0$. The outer expansion is an ordinary power series in integer powers of $\epsilon$, like (3.3), with leading term

$$c^0(x) = \frac{f}{\varphi'(x)}.$$ 

The inner limit ($x \to 0$) of $c^0(x)$ is

$$c^0(x) = \frac{f}{\varphi''(0)x} - \frac{f \cdot \varphi'''(0)}{2 \cdot (\varphi''(0))^2} + O(x).$$ 

(3.15)

Since $c^0(x)$ diverges like $\frac{1}{x}$ as $x \to 0$, we anticipate that there is a boundary layer at $x = 0$ in which $c(x, \epsilon)$ is much greater than unity. The inner expansion cannot be an ordinary power series starting with an $O(1)$ term. In problem 1.2 we showed that dominant balance in the ODE (3.2) leads to the form of inner solution,

$$c(x, \epsilon) = \frac{1}{\sqrt{\epsilon}} C \left( X := \frac{x}{\sqrt{\epsilon}} \right).$$
Chapter 3. Matched asymptotic expansions

The ODE for $C(X, \epsilon)$ is

$$C_X + \frac{\varphi'(\sqrt{\epsilon}X)}{\sqrt{\epsilon}} C = f.$$  

It appears that the correct asymptotic sequence for the inner expansion is $\epsilon^{-\frac{1}{2}}, 1, \epsilon^{\frac{1}{2}}, \ldots$ and the inner expansion takes the form

$$c \sim \frac{1}{\sqrt{\epsilon}} C^0 + C^1 + \sqrt{\epsilon} C^2 + \ldots.$$  

How many terms of inner expansion are needed to asymptotically match the inner limit (3.15) of leading order outer expansion? Since (3.15) contains the constant $-\frac{f}{2 (\varphi''(0))^2}$ we anticipate we’ll need $C^0$ and the $O(1)$ term $C^1$. Hence we construct the two term inner expansion $\frac{1}{\sqrt{\epsilon}} C^0(X) + C^1(X)$. The perturbation equations for $C^0$ and $C^1$ are

$$C^0_X + \varphi''(0)X C^0 = f,$$
$$C^1_X + \varphi''(0)X C^1 = -\frac{\varphi'''(0)}{2} X^2 C^0. \quad (3.16)$$

Integral representations of solutions for $C^0$ and $C^1$ which vanish at $X = 0$ are easily constructed. In fact the construction of $C^0$ was carried out in problem 1.2. Asymptotic evaluation of these integral representations as $X \to \infty$ gives the outer limits of $C^0(X)$ and $C^1(X)$. Here, we do a shortcut: The outer limits have components with the exponential factor $e^{-\frac{\varphi''(0)}{2}X^2}$, and these are exponentially small in any intermediate limit with $p < \frac{1}{2}$: We have $X^2 = \frac{x^2}{\epsilon}$, and for $x = O(\epsilon^p)$, $X^2 = O(\epsilon^{2p-1}) \gg 1$ if $p < \frac{1}{2}$. Aside from the exponentially small terms, the outer limits of $C^0$ and $C^1$ are power series in $\frac{1}{X}$, and coefficients of these series are determined by substitution into (3.16).

In this way we find

$$C^0(X) = \frac{f}{\varphi''(0)X} + O\left(\frac{1}{X^3}\right),$$
$$C^1(X) = -\frac{f}{2} \frac{\varphi'''(0)}{(\varphi''(0))^2} + O\left(\frac{1}{X^2}\right).$$

Notice that the leading term of $C^1(X)$ is precisely the constant that appears in (3.15). The Emperor has foreseen this. To summarize, the outer ($X \to \infty$)
limit of two-term inner expansion written in terms of $x$ is

$$
\frac{1}{\sqrt{\epsilon}} C^0 \left( \frac{x}{\sqrt{\epsilon}} \right) + C^1 \left( \frac{x}{\sqrt{\epsilon}} \right) = \frac{f}{\varphi'(0)x} - \frac{f''(0)}{2 (\varphi'(0))^2} + O \left( \frac{\epsilon}{x^3} + \frac{\epsilon}{x^2} \right). \quad (3.17)
$$

We are now ready to determine the overlap domain of the leading order outer expansion $\epsilon^0(x)$ with the two-term inner expansion $\frac{1}{\sqrt{\epsilon}} C^0(X) + C^1(X)$. First recall $p < \frac{1}{2}$ is a requirement for the validity of (3.17). Next the difference between (3.15), (3.17) is $O \left( \epsilon^x x^3 + \epsilon^y x^2 \right)$, and this is $o(1)$ in any intermediate limit $x = O(\epsilon^p)$ with $\epsilon^{1-5p} + \epsilon^p = o(1)$, so $0 < p < \frac{1}{3}$. These $p$'s have $p < \frac{1}{2}$ automatically. In summary, the overlap domain is $x = O(\epsilon^p)$, with $0 < p < \frac{1}{3}$.

**Matched asymptotic expansions in practice are practice**

Constructing matched asymptotic expansions of solutions to real world problems is “art” instead of “method”. Here is a preview of the “ocean (of real singular perturbation problems) as seen from space”.

Suppose the domain of independent variables is a region $R$ in $\mathbb{R}^n$. Let $S$ denote the set of “singular points” in $\hat{R}$ (the closure of $R$), so the outer solution is non-uniformly valid as you approach any point of $S$. $S$ is not always confined to a portion of the boundary $\partial R$. If $S$ is an $n-1$-dimensional surface in the interior, the neighborhood of this surface where the inner solution lives is called an *internal layer* instead of a boundary layer. $S$ may be “smaller” than an $n-1$-dimensional surface: For a vortex filament in a three-dimensional fluid flow, $S$ is a one-dimensional curve in $\mathbb{R}^3$. There is also the “degree of non-uniformity”: The outer solution may in fact converge uniformly, but some derivative does not. We refer to such structures in the asymptotic solution as “derivative” or “corner” layers. The preceding bits of nomenclature describe commonplace occurrences but are not exhaustive. You just can’t imagine all the kinds of singularity that Great Nature will throw at you.

The few simple ideas you have, like “matched asymptotic expansions” are not “machines” so you flip the switch and “let the math be smart”. Rather, they are guides to a *practice* whose details are improvised as you go along. For instance, suppose dimensional analysis gives you widely separated time or space scales, so you think: “Outer expansion for large scales and inner for small”. The physical character of problem helps identify the singular set $S$ of outer solution. For example, a shock layer in a gas lives around a surface, a vortex filament around a curve, and a “monopole” in some nightmarish
modern field theory around a point. Dominant balances help decide the
asymptotic sequences of the inner and outer expansions. Singular behaviors
of solutions to reduced equations also have their say: The log \( r \) associated
with the two-dimensional Laplacian often dooms you to log \( \epsilon \). Next, how
many terms of inner and outer expansions are necessary to achieve a given
order of matching? A priori you \textit{just don’t know}. Wade in and be prepared
for a fight. Determine by trial and (mostly) error what works.

Finally and most important: You usually construct candidate inner and
outer expansions without a-priori knowledge of any exact solution, and one
or the other or both are not unique. It falls to asymptotic matching as an
assumed guiding principle to select the constants of integration that \textit{cannot}
be determined within one or the other limit process alone. Sometimes you can’t
make it (asymptotic matching) happen. Maybe there is another distinguished
limit on a scale between your proposed inner and outer limits which you
missed on the first try, and what you are really up against is a “triple layer
cake” with two matching processes.

To summarize: Singular perturbation is “practice” first and “field of
knowledge” second. May the Force be with you. You will need it.

Basic tutorial example

Boundary layers, internal layers and derivative layers all appear in a fa-
mous example due to Cole and Kevorkian. We analyze the \( \epsilon \to 0 \) limits of
\( y(x, \epsilon) \) which satisfy ODE

\[
\epsilon y'' + yy' - y = 0 \quad (3.18)
\]
in \( x > 0 \), and

\[
y(0, \epsilon) = a, \quad (3.19)
\]
where \( a \) is a constant independent of \( \epsilon \). The solutions of the outer reduced
equation \( yy' - y = 0 \) and \( y \equiv 0 \) or \( y = x + c \) for any constant \( c \). These
outer solutions are also solutions of the full ODE (3.18), and we will discuss
solutions of (3.18), (3.19) that exist on \([0, \infty)\), and asymptote to one of these
as \( x \to +\infty \).

\textsuperscript{2}Cole and Kevorkian do a standard two-point boundary value problem with prescribed
values of \( y \) at \( x = 0,1 \). The class of solutions we examine here requires less elaboration
of special cases, but displays \textit{all} the interesting structures: boundary, internal and corner
layers.
In view of the \( \epsilon \)-independent boundary condition (3.19), we examine the inner limit process \( \epsilon \rightarrow 0 \) which balances the \( \epsilon y'' \) and \( yy' \) terms of (3.18) with \( y = O(1) \). The appropriate representations of \( y \) are

\[
y = Y \left( X := \frac{x - x^*}{\epsilon}, \epsilon \right).
\] (3.20)

We assume that \( Y^0(X) := \lim_{\epsilon \rightarrow 0} Y(X, \epsilon) \) exists, and \( x^* \) is an arbitrary constant (as allowed by the autonomy of the ODE (3.18)). The choice \( x^* = 0 \) is appropriate for a boundary layer at \( x = 0 \), but we will discuss the possibility of internal layers about various \( x^* > 0 \). The ODE (3.18) in terms of \( Y(X, \epsilon) \) is

\[
Y'' + YY' - \epsilon Y = 0 \tag{3.21}
\]

where \( Y' := Y_X(X, \epsilon) \). The leading order inner solution \( Y^0(X) \) satisfies the reduced equation

\[
(Y^0)' + Y^0(Y^0)' = 0. \tag{3.22}
\]

There is an elementary first integral: For any solution \( Y^0 \) of (3.22)

\[
E^0(Y^0, (Y^0)') := (Y^0)' + \frac{(Y^0)^2}{2} \tag{3.23}
\]

is a constant independent of \( X \). The phase plane of ODE (3.22), consisting of the level curves of \( E^0 \) in the \( Y^0, (Y^0)' \) plane oriented by increasing \( X \), is depicted in Figure 3.3. The curves are the parabolas \( (Y^0)' = E - \frac{(Y^0)^2}{2} \) and orientations are clear: In \( (Y^0)' > 0 \), we have “flow to the right” consistent with \( Y^0 \) increasing in \( X \), and in \( (Y^0)' < 0 \), “flow to the left”. Notice that the whole \( Y^0 \) axis is a line of critical points.

The phase plane naturally divides into four regions (I–IV in Figure 3.3). The “I” region contains the portions of \( E > 0 \) trajectories in \( (Y^0)' > 0 \). The corresponding solutions for \( Y^0(X) \) are

\[
Y^0(X) = \sqrt{2E} \tanh \sqrt{\frac{E}{2}} (X - X_*). \tag{3.24}
\]

Here, the arbitrary constant \( X_* \) arises from the autonomy of the ODE (3.22). We could absorb \( X_* \) as an \( \epsilon \)-term of \( x_* \), which appears in the definition of the inner variable \( X \) in (3.20). The main point of the “I” solutions is that they asymptote to \( \pm \sqrt{2E} \) as \( X \rightarrow \pm \infty \), and \( (Y^0, (Y^0)') \) “starts” from the
The "II" solutions are given by

\[ Y^0(X) = \frac{\sqrt{2E}}{\tanh \sqrt{\frac{E}{2}(X - X_*)}}, \quad (3.25) \]

in \( X > X_* \). Here, \( E > 0 \) and \( X_* \) are constants. These "start" from a vertical asymptote at \( X = X_* \) and converge to \( \sqrt{2E} \) as \( X \to \infty \), so the \((Y^0, (Y^0)')\) trajectory "falls onto" \((\sqrt{2E}, 0)\) as \( X \to \infty \). The darkened trajectory which falls onto \((0, 0)\) is the limit of "II" trajectories as \( E \to 0 \). The corresponding \( Y^0(X) \) is the \( E \to 0 \) limit of (3.25) computed by L'Hospital's rule,

\[ Y^0(X) \to \frac{2}{X - X_*} \quad (3.26) \]

as \( E \to 0 \). Region "III" solutions can be obtained by replacing \( Y \) and \( X \) in a "II" solution by \(-Y\) and \(-X\). These solutions "start" from \(-\sqrt{2E}\)
as $X \to -\infty$ and develop a vertical asymptote at some $X$. They do not correspond to a solution $y(x, \epsilon)$ to (3.18) which lives on $[0, \infty)$. Neither do “IV” solutions which begin and end with vertical asymptotes.

We now use the outer solutions $y \equiv 0$ or $y = x + C$, and type I or II inner solutions to construct asymptotic solutions $y(x, \epsilon)$ of (3.18), (3.19) on $[0, \infty)$. We’ll start with the subclass of solutions with a boundary layer at $x = 0$ and one of the outer solutions applies as $\epsilon \to 0$ with $x > 0$ fixed. Both type I and II solutions (3.24), (3.25) have $E > 0$ and asymptote to the positive constant $\sqrt{2E}$ as $X \to \infty$. Hence the leading order outer solution which matches one of the inner solutions (3.24) or (3.25) is $y^0 = x + c$ with $c = \sqrt{2E}$. It is convenient to parametrize the inner solutions with $c > 0$ instead of $E$, so we write the type I and II inner solutions as

I) $Y^0(X) = c \tanh \frac{c}{2}(X - X_*)$,  
II) $Y^0(X) = \frac{c}{\tanh \frac{c}{2}(X - X_*)}$. 

The inner solution satisfies the boundary condition $Y^0(0) = a$. If $|a| < c$, the “I” solution with

$$X_* = -\frac{2}{c} \text{arctanh} \frac{a}{c}$$

satisfies $Y^0(0) = a$. If $a > c$, then we get the “II” solution with

$$X_* = -\frac{2}{c} \text{arctanh} \frac{c}{a}.$$ 

Figure 3.4 shows the inner solution in relation to the outer solution and exact solution as $a$ decreases from above $c > 0$ to $-c$. 


Figure 3.4
Chapter 3. Matched asymptotic expansions

The cases $a > 0$ yield “garden variety boundary layers”. But notice that as $a$ approaches $-c$ from above, it appears that the “boundary layer detaches from $x = 0$”.

This brings up the next case, $a < -c < 0$, depicted in Figure 3.5. For $a < -c < 0$, there is no type I or II solution with $Y^0(0) = a$ which asymptotes to $c > 0$ as $X \to \infty$. Instead we get $y^0 = x + a$ as an outer solution in $0 < x < x_* := -\frac{a}{2} - \frac{c}{2}$. At $x = x_*$, the “left” and right outer solutions $x + a$ and $x \to c$ are opposite and equal, and this is where we can fit an internal layer, with inner solution

$$Y^0 \left( X := \frac{x - x_*}{\epsilon} \right) = \frac{c - a}{2} \tanh \left( \frac{c - a}{4} X \right).$$

To summarize: For any $c > 0$, and any $a$, we have constructed an asymptotic solution $y(x, \epsilon)$ of ODE (3.18) with $y(0, \epsilon) = a$ and $y(x, \epsilon) \sim x + c$ as $x \to \infty$.

Next, we analyze solutions that asymptote to zero or to $x - c$ with $c > 0$ as $x \to \infty$. The $\epsilon \to 0$ limit processes (inner or outer) with $y = O(1)$ are not sufficient to deal with them. Here is the first indication of trouble: The $E \to 0$ limit (3.26) of “II” solutions asymptotes to zero as $X \to \infty$, and if
we set $X_* = -\frac{2}{a}$ for any $a > 0$, we get $Y^0(X) = \frac{a}{1 + \frac{2X}{a}}$ which satisfies the boundary condition $Y^0(0) = a$. This inner solution exhibits leading order matching with $y^0 \equiv 0$ but the algebraic decay like $\frac{1}{X}$ as $X \to \infty$ is wrong. The linearization of the full ODE (3.18) about $y \equiv 0$ is $\epsilon y'' - y = 0$, and the solutions which decay to zero as $x \to \infty$ are proportional to $e^{-\frac{1}{\sqrt{\epsilon}}}$.

The length constant $\sqrt{\epsilon}$ of this exponential decay is longer than the assumed characteristic length $\epsilon$ of the “old” inner limit (3.20) so of course the old inner limit does not resolve it. What actually happens: There is an $\epsilon \to 0$ distinguished limit of ODE (3.18), $y = O(\sqrt{\epsilon})$, $x - x_* = O(\sqrt{\epsilon})$, which balances all three terms: Represent $y(x, \epsilon)$ by

$$y = \sqrt{\epsilon} z \left( s := \frac{x - x_*}{\sqrt{\epsilon}}, \epsilon \right).$$

The ODE for $z(s, \epsilon)$ is

$$z'' + zz' - z = 0,$$  \hspace{1cm} (3.29)

where $z' := z_s$. For any solution $z$, the first integral

$$E(z, z') := z' + \log |z' - 1| + \frac{z^2}{2}$$  \hspace{1cm} (3.30)

is a uniform constant independent of $s$. Figure 3.6 shows the phase plane of ODE (3.29), consisting of level curves of $E(z, z')$ oriented by increasing $s$. 


Figure 3.6
Chapter 3. Matched asymptotic expansions

The solution with \( z(0, \epsilon) = \frac{a}{\sqrt{\epsilon}} > 0 \) and \( y(\infty, 0) = 0 \) generates the portion of the incoming separatrix \( C \) between \( z = \frac{a}{\sqrt{\epsilon}} \) and \( z = 0 \). For \(|z|, |z'| \gg 1\), the log term in the first integral (3.30) is asymptotically small compared to the remaining terms \( z' \) and \( \frac{z^2}{2} \), and we’re back to the reduced first integral in (3.23). Its no surprise that the asymptotic behavior of \( y(x, \epsilon) = \sqrt{\epsilon} z\left(\frac{x}{\sqrt{\epsilon}}, \epsilon\right) \) as \( \epsilon \to 0 \) with \( X := \frac{x}{\epsilon} \) fixed is going to be the previous “naive” inner solution \( Y^0(X) = \frac{a}{1 + \frac{X^2}{2}} \). As \( s \to \infty \), \( z(s, \epsilon) \to 0 \) so the linearization of (3.29) applies, and the exponential decay \( e^{-s} = e^{-\frac{s}{\sqrt{\epsilon}}} \) follows.

In general, the trajectories in Figure 3.6 for \(|z|, |z'| \gg 1\) have a nice correspondence with the parabolas in the “reduced” phase plane of Figure 3.3. In particular, the dotted curve in Figure 3.6 is the parabola \( E = z' + \frac{z^2}{2} \) for some large \( E \), and we see clearly how the exact trajectories “hug” it. In the strip where \( z' - 1 = O(1) \), we see a fundamental change from the reduced phase plane. The critical points along the horizontal axis of the “old” phase plane have been “fired” under the new management, and only the origin \((z, z') = (0, 0)\) survives, as a saddle point. We’ve indicated regions I and II in the new phase plane which are clear counterparts of regions I, II in the old phase plane. Notice that the I and II trajectories asymptote to the horizontal line \( z' = 1 \) as \( s \to \infty \), instead of just “stopping” at some point along a line of critical points. For instance, the trajectory \( aa' \) corresponds to a solution with an internal layer, and the fact that it asymptotes to \( z' = 1 \) as \( s \to +\infty \) or \(-\infty \) indicates its matching with outer solutions with \( y' = z' = 1 \) away from the internal layer.

There is still more information in Figure 3.6 that was hidden from view in the \( \epsilon \to 0 \), \( y = O(1) \) limits. In particular, we can analyze solutions \( y(x, \epsilon) \) that asymptote to \( x - c \) as \( x \to \infty \) with \( c > 0 \). For instance, consider the trajectory \( bb' \) that starts out by “hugging” the separatrix \( C \). It comes close to the saddle point \((0, 0)\), “loitering” there for a long interval of \( s \), and finally “hugs” \( z' = 1 \) as \( s \to +\infty \). It is not hard to visualize the corresponding solution for \( y(x, \epsilon) \), as depicted in Figure 3.7.
The dotted curve is the portion of separatrix solution that starts from $y = a$ at $x = 0$ and asymptotes to zero as $x \to \infty$. Notice that $y = y(x, \epsilon)$ has a derivative layer at $x = c$. The structure of $y(x, \epsilon)$ in the derivative layer is dictated by the outgoing separatrix marked $C'$ in Figure 3.6. The solution for $z$ corresponding to the separatrix $C'$ has $z \to 0$ as $s \to -\infty$, $z' \to 1$ as $s \to +\infty$. The corresponding $y(x, \epsilon)$ has $y \to 0$ as $\frac{x}{\sqrt{\epsilon}} \to -\infty$ and $y' \to 1$ as $\frac{x}{\sqrt{\epsilon}} \to +\infty$. The thickness of the derivative layer is $O(\sqrt{\epsilon})$. Finally, there are “two-derivative layer” solutions with $y(0, \epsilon) = a < 0$ and $y \sim x - c$, $c > a$, as $x \to +\infty$, as depicted in Figure 3.8.
Figure 3.8

The reader will identify the corresponding trajectories in the $z, z'$ phase plane.
Chapter 4

Matched asymptotic expansions in PDE

We construct matched asymptotic expansions to solutions of PDE. A key technical idea arises: We’ve mentioned that candidate inner and outer expansions obtained by direct solution of inner and outer equations may be non-unique, and in some cases, the non-uniqueness is resolved by matching. Yes, it is so. But more significant is how? At the heart of these problems there often lies an inhomogeneous, linear perturbation equation. The linear operator has a non-trivial null space, and it is the Fredholm solvability conditions which “lay down the law”, resolving the previously undetermined parameters in inner and outer expansions. In this way we answer questions like: “Where do the internal layers go?” Or: “How do they move?”, or “How are eigenvalues of Laplacian with zero boundary condition perturbed by ‘punching small holes’ in the domain?”

Moving internal layers

We present a classic example of moving internal layers in solutions of a non-linear diffusion PDE. Here is a specific physical context: A sheet of liquid crystal occupies the $x, y$ plane. Its molecular units are rod-like polymers as depicted in Figure 4.1. The state variable quantifying configurations is the angle $\theta(x,t)$ of rods at position $x = (x, y)$ and time $t$, as indicated in Figure 4.1. These rods have no head-tail directionality, so rods with angles $\theta$ and $\theta + \pi$ are physically indistinguishable. The energy of a given configuration has these two components: (i) There is elastic resistance against spatial variations of angulation, modeled by an elastic energy per unit area propor-
tional to $|\nabla \theta|^2$. (ii) We subject the liquid crystal to a uniform electric field in the $x^1$ direction and an induced polarization in the rod-shaped molecules tends to align them parallel to the $x$ axis. Assuming that the given uniform electric field is much stronger than the electric field due to polarization, there is an electrical energy per unit area proportional to $\sin^2 \theta$. In summary, the (dimensionless) energy of a given configuration $\theta(x, t)$ is

$$U = \int_{\mathbb{R}^2} \left\{ \frac{1}{2} |\nabla \theta|^2 + E^2 \sin^2 \theta \right\} \, dx. \quad (4.1)$$

Here, $E$ is proportional to the electric field strength. The (dimensionless) time dynamics of $\theta(x, t)$ is the gradient flow of this energy,

$$\theta_t = -\frac{\delta U}{\delta \theta} = \Delta \theta - E^2 \sin 2\theta,$$

or setting $\epsilon = \frac{1}{E}$,

$$\epsilon^2 \theta_t = \epsilon^2 \Delta \theta - \sin 2\theta. \quad (4.2)$$

We analyze a specific class of solutions to (4.2) in the strong electric field limit $\epsilon \to 0$. There are obvious $\epsilon = 0$ outer solutions $\theta \equiv \pi n$, $n =$ integer, which are also energy minimizing exact solutions of the full equation for $\epsilon > 0$. For any $n$, $\theta \equiv \pi n$ represents the same physical configuration of all molecules aligned parallel to the $x$ axis. Next, there are time independent solutions with internal layers in which $\theta$ increases from $\pi n$ to $\pi(n + 1)$ as you cross the layer. Let $e$ be any unit vector, so $s = x \cdot e$ represents displacement in the $e$ direction. $\theta = \vartheta(s)$ is a time independent solution of (4.2) if $\vartheta(s)$ satisfies ODE

$$\epsilon^2 \vartheta'' - \sin 2\vartheta = 0. \quad (4.3)$$
For any solution $\vartheta(s)$ of (4.3), the first integral

$$E(\vartheta, \vartheta') =: \frac{\epsilon^2}{2} (\vartheta')^2 - \sin^2 \vartheta$$

is a uniform constant. Figure 4.2 depicts the phase plane, consisting of level curves of $E$ in the $(\vartheta, \vartheta')$ plane, oriented by increasing $s$.

The heteroclinic trajectories connecting neighboring saddle points all have $E = 0$. In particular, the trajectory $\vartheta' > 0$ connecting $(0, 0)$ and $(\pi, 0)$ has

$$\epsilon \vartheta' = \sqrt{2} \sin \vartheta$$

and the solution of this first order ODE with $\vartheta(0) = \frac{\pi}{2}$ is

$$\vartheta = 2 \arctan e^{\frac{\sqrt{2} s}{\epsilon}}.$$  \hspace{1cm} (4.5)

Solutions corresponding to any other heteroclinic trajectory in Figure 4.2 can be constructed by applying obvious symmetry operations to (4.6). Figure 4.3 visualizes the liquid crystal configuration corresponding to $\theta = \vartheta(s)$ in (4.6). The main point is that the rods rotate from $\theta = 0$ to $\theta = \pi$ as we cross an internal layer of thickness $\epsilon$ about $s = 0$.

We expect that internal layers with local structure like Figure 4.3 arise naturally from rather general initial conditions. For instance, suppose we
have initial data with $\theta \to 0$ as $x \to -\infty$ and $\theta \to \pi n$ as $x \to +\infty$. As we go along any curve from $x = -\infty$ to $x = +\infty$, we see the rod orientation make $n \pi$ rotations. This topological feature of the liquid crystal configuration survives as the angle field $\theta(x, t)$ evolves under the time dynamics (4.2). We expect that the liquid crystal sheet “anneals” into domains within which rods are close to the energy minimizing horizontal orientation, and the rotation of rod orientation is compressed into $n$ internal layers of thickness $\epsilon$, each accounting for one “$\pi$-rotation”. The balance between $\epsilon^2 \theta_t$ and $\sin 2\theta$ terms in (4.2) suggests that the characteristic time of this “initial annealing” is $\epsilon^2$. Here we assume that the initial annealing has already happened, so $\mathbb{R}^2$ is divided up into domains of horizontally oriented rods, separated by internal layers of thickness $\epsilon$ where “$\pi$-rotations” happen. We want to analyze the asymptotic ($\epsilon \to 0$) time dynamics of these internal layers, idealized as geometric curves in $\mathbb{R}^2$.

Figure 4.4 shows the setup. The curve $C$ is the “centerline” of the layer, where $\theta = \frac{\pi}{2}$. We want to figure out the normal velocity of $C$ when it crosses some fixed point $p$. Due to the autonomy of the PDE (4.2), the time dynamics of $C$ is expected to be autonomous as well, so we can set the time of crossing to zero with no loss of generality. It is convenient to shift the origin of cartesian coordinates to $p$, and orient the axes $\hat{x}$ and $\hat{y}$ so that $\hat{x}$ is the tangent of $C$ at $p$ when it crosses, and $\hat{y}$ is oriented in the direction of increasing $\theta$. The local description of $C$ near $(x, y, t) = (0, 0, 0)$ is

$$y = h(x, t, \epsilon),$$

where $h(0, 0, \epsilon) = 0$ and $h_x(0, 0, \epsilon) = 0$ as follows from the origin and ori-
entation of the $x,y$ cartesian coordinates. The normal velocity we want to compute is

$$v := h_t(0,0,\epsilon).$$

Since the internal layer has $0(\epsilon)$ thickness about the curve (4.7), we introduce the translated and scaled $y$-displacement

$$Y := \frac{y - h(x,t,\epsilon)}{\epsilon}$$

and the inner solution has representation

$$\theta = \Theta(x,Y,t,\epsilon).$$

Now it is crunch time, to substitute (4.10) into PDE (4.2). We do the usual chain rule calculation and it starts out looking awful:

$$\epsilon^2 \Theta_t - \epsilon h_t \Theta_y = \epsilon^2 \Theta_{xx} - 2\epsilon h_x \Theta_{xy} - \epsilon \Theta_y h_{xx} + (1 + h_x^2) \Theta_{yy} - \sin 2\Theta.$$

But now we focus specifically on the “crossing event” with $x = 0$, $t = 0$. Since $h_x(0,0,\epsilon) = 0$ many obnoxious terms are rubbed out, leaving

$$\Theta_{yy} - \sin 2\Theta = \epsilon(h_{xx} - h_t)\Theta_y + \epsilon^2(\Theta_t - \Theta_{xx})$$  \hspace{1cm} (4.11)
Chapter 4. Matched asymptotic expansions in PDE

at \( x = 0, t = 0 \).

We analyze the two-term inner expansion

\[
\Theta(x = 0, Y, t = 0, \epsilon) \sim \Theta^0(Y) + \epsilon \Theta^1(Y).
\]  

(4.12)

The perturbation equations for \( \Theta^0 \) and \( \Theta^1 \) are

\[
\Theta^0_{YY} - \sin 2\Theta^0 = 0, \tag{4.13}
\]

\[
\Theta^1_{YY} - 2 \cos 2\Theta^0 \Theta^1 = (h^0_{xx} - h^0_t)(0, 0)\Theta^0_Y. \tag{4.14}
\]

Here, \( h^0 := h(x, t, \epsilon = 0) \). Effective boundary conditions on \( \Theta^0 \) and \( \Theta^1 \) at \( y = \infty \) follow from matching with outer solutions: The outer solutions in \( y > h^0(x, t) \) and \( y < h^0(x, t) \) to all orders in \( \epsilon \) are \( \theta \equiv 0 \) and \( \theta \equiv \pi \), respectively. Hence we have \( \Theta^0 \to 0 \) as \( y \to -\infty \), and \( \Theta^0 \to \pi \) as \( y \to +\infty \), and \( \Theta^1 \to 0 \) as \( |y| \to \infty \). The solution for \( \Theta^0 \) is just (4.6) with \( \frac{s}{\epsilon} \) replaced by \( Y \),

\[
\Theta^0 = 2 \arctan(e^{\sqrt{2}Y}). \tag{4.15}
\]

But now a question: We’re supposed to compute the leading approximation \( h^0_t \) to the normal velocity. We have no equation that explicitly gives \( h^0_t \). All we apparently have is the appearance of \( h^0_t \) in the right-hand side of the linear inhomogeneous ODE (4.14) for \( \Theta^1 \). So where is this determination of \( h^0_t \) coming from, anyway?

In essence, from simple linear algebra: First, the variational operator associated with the non-linear ODE (4.13) is

\[
(LV)(Y) := V''(Y) - 2 \cos 2\Theta^0(Y)V(Y). \tag{4.16}
\]

The homogeneous variational boundary value problem, \((LV)(Y) = 0\) with \( V(Y) \to 0 \) as \( |Y| \to \infty \) has solutions proportional to \( V(Y) = \Theta^0_Y \), as can be seen by differentiating (4.13) with respect to \( Y \). Next, \( \Theta^1 \) satisfies the inhomogeneous variational boundary value problem,

\[
(L\Theta^1)(Y) = F(Y) := (h^0_t - h^0_{xx})\Theta^0_Y(Y) \tag{4.17}
\]

with \( \Theta^1(Y) \to 0 \) as \( |Y| \to \infty \). Since the homogeneous problem has a non-trivial solution, the inhomogeneous problem has a solvability requirement which can be derived in the usual linear algebra way: Take the inner product of (4.16) with a test function \( V \),

\[
(V, L\Theta^1) := \int_{-\infty}^{\infty} VL\Theta^1 dx = (V, F). \tag{4.18}
\]
Chapter 4. Matched asymptotic expansions in PDE

$L$ is self-adjoint, meaning $(V, LW) = (LV, W)$ for all $V, W$ that vanish as $|Y| \to \infty$. The proof is two integrations by parts. Hence, (4.18) becomes

$$(LV, \Theta^1) = (V, F),$$

and it follows that the boundary value problem for $\Theta^1$ is solvable only if $(V, F) = 0$ for all solutions $V$ of the homogeneous variational boundary value problem. In particular,

$$0 = (\Theta^0_Y, F) = (h^0_t - h^0_{xx})(0, 0) \int_{-\infty}^{\infty} (\Theta^0_Y)^2 \, dy$$

and the solvability condition is

$$h^0_t(0, 0) = h^0_{xx}(0, 0). \quad (4.19)$$

Here, $h^0_t(0, 0)$ is the leading approximation to the normal velocity of centerline $C$ as it crosses $p$ in Figure 4.4, and $h^0_{xx}(0, 0)$ is the curvature $\kappa$ of $C$ at $p$. Hence, to leading approximation, the centerline $C$ evolves in time according to the geometrical law

normal velocity $\sim v^0 = \kappa = \text{curvature}. \quad (4.20)$

**Asymptotic Lagrangian**

We modify the PDE (4.2) into a non-linear wave equation

$$\epsilon^2 \theta_{tt} = \epsilon^2 \Delta \theta - \sin 2\theta, \quad (4.21)$$

which derives from the Lagrangian density

$$L := \frac{\epsilon^2}{2} (\theta_t^2 - |\nabla \theta|^2) - \sin^2 \theta. \quad (4.22)$$

We examine $\epsilon \to 0$ solutions so that $\theta \to \pi$ in the exterior of some world tube $D$ in $2 + 1$ dimensional spacetime, and $\theta \to 0$ in the interior, and there is an internal layer of thickness $\epsilon$ centered about the two-dimensional timelike surface $\partial D$ where $\theta = \frac{\pi}{2}$. This is depicted in Figure 4.5. The cross-section of $\partial D$ at any fixed $t$ is a simple closed curve, and we analyze the possible time evolutions.

Let $p$ be a point on $\partial D$. As before, we take $p$ as the origin of the $(x, y, t)$ spacetime coordinates. This means that the cross-section of $\partial D$ at time $t$
the “front”) crosses \((x, y) = (0, 0)\) at time \(t = 0\). If we represent the front by \(y = h(x, t, s)\), this means \(h(0, 0, \epsilon) = 0\). As further refinement might be to adopt spacetime coordinates \((x, y, t)\) of a reference frame whose \(x\) axis is parallel to the \(t = 0\) front at \((x, y) = (0, 0)\), and the normal velocity of the front at \((x, y, t) = (0, 0, 0)\) vanishes. In this “instantaneous rest frame” we’d have \(h_x(0, 0, \epsilon) = 0\) and \(h_t(0, 0, \epsilon) = 0\). But this time, we’ll forego the instantaneous rest frame. We find that the “direct” derivation of the PDE for \(h^0(x, t) := h(x, t, \epsilon = 0)\) is not so hard after all. The final act is to puzzle out its spacetime geometric meaning.

The ansatz for \(\theta\) in the internal layer is as before,
\[
\theta = \Theta \left( x, Y := \frac{y - h(x, t, \epsilon)}{\epsilon}, t, \epsilon \right).
\]
What is different is that it goes into the wave equation (4.21): We get
\[
\epsilon^2 \Theta_{tt} - 2\epsilon h_t \Theta_{tY} - \epsilon h_{tt} \Theta_Y + h_t^2 \Theta_{YY} \\
= \epsilon^2 \Theta_{xx} - 2\epsilon h_x \Theta_{xY} - \epsilon h_{xx} \Theta_Y + h_x^2 \Theta_{YY} + \Theta_{YY} - \sin 2\Theta = 0,
\]
or
\[
(1 + h_x^2 - h_t^2)\Theta_{YY} - \sin 2\Theta \\
= \epsilon \{(h_{xx} - h_{tt})\Theta_Y + 2(h_x \Theta_{xY} - h_t \Theta_{tY})\} + \epsilon^2 (\Theta_{tt} - \Theta_{xx}).
\]
Now substitute two-term expansions
\[
\Theta \sim \Theta^0 + \epsilon \Theta^1, \quad h \sim h^0 + \epsilon h^1
\]
to obtain perturbation equations

\[(1 + h_0^0 - h_1^0) \Theta_{YY}^0 - \sin 2\Theta^0 = 0, \tag{4.25}\]

\[
(1 + h_0^0 - h_1^0) \Theta_{YY}^1 - 2(\cos 2\Theta^0) \Theta^1 \\
= (h_{xx}^0 - h_{tt}^0) \Theta_Y^0 + 2(h_x^0 \Theta_{xY}^0 - h_t^0 \Theta_{tY}^0) \\
+ 2(h_t^0 h_1^1 - h_x^0 h_x^1) \Theta_{YY}^0. \tag{4.26}\]

The effective matching boundary conditions are as before, \(\Theta^0 \to 0\) as \(Y \to -\infty\), \(\Theta^0 \to \pi\) as \(Y \to +\infty\), and \(\Theta^1 \to 0\) as \(|Y| \to \infty\). The solution for \(\Theta^0\) is

\[\Theta^0 = 2 \arctan \sqrt{\frac{2}{1 + h_x^0 - h_t^0}} Y.\tag{4.27}\]

Again, solvability of the boundary value problem for \(\Theta^1\) requires that the right-hand side of (4.26) be orthogonal to \(\Theta_Y^0\), and this gets nicely expressed as

\[
\partial_x \left( h_x^0 \int_{-\infty}^{\infty} \Theta_Y^0 dy \right) - \partial_t \left( h_t^0 \int_{-\infty}^{\infty} \Theta_Y^0 dy \right) = 0. \tag{4.28}\]

Notice that \(\int_{-\infty}^{\infty} \Theta_Y^0 \Theta_Y^0 dy = 0\), so the obnoxious little term with \(h_t^0 h_1^1 - h_x^0 h_x^1\) is history. In (4.28), we need to evaluate \(\int_{-\infty}^{\infty} \Theta_Y^0 dy\). We could do it using the solution (4.27) for \(\Theta^0\). But here is a smarter method that requires only knowledge of the heteroclinic orbit is the \(\Theta^0, \Theta_Y^0\) phase plane: The first integral of the ODE (4.26) with \(\Theta^0 \to 0\) as \(Y \to -\infty\) and \(\Theta^0 \to \pi\) as \(Y \to \infty\) is

\[\Theta_Y^0 = \sqrt{\frac{2}{1 + h_x^0 - h_t^0}} \sin \Theta.\tag{4.29}\]

The graph of \(\Theta_Y^0\) versus \(\Theta^0\) in \(0 \leq \Theta^0 \leq \pi\) is the heteroclinic orbit referred to before. We now calculate with the help of (4.29),

\[
\int_{-\infty}^{\infty} \Theta_Y^0 \, dY = \sqrt{\frac{2}{1 + h_x^0 - h_t^0}} \int_0^\pi \sin \Theta^0 \, d\Theta^0 = 2 \sqrt{\frac{2}{1 + h_x^0 - h_t^0}}.
\]

(4.28) now becomes an explicit PDE for \(h^0(x,t)\):

\[
\partial_t \left( \frac{h_t^0}{\sqrt{1 + h_x^0 - h_t^0}} \right) - \partial_x \left( \frac{h_x^0}{\sqrt{1 + h_x^0 - h_t^0}} \right) = 0. \tag{4.30}\]
This PDE is shouting out that it is a Lagrangian dynamics, with Lagrangian density

\[ L^0 := \sqrt{1 + h_0^2 - h_t^2}. \]  

(4.31)

What is this beast? In sophomore vector calculus you learn that the surface area of the graph \( z = h(x, y) \) above the rectangle \([x, x + dx] \times [y, y + dy]\) is \( da = \sqrt{1 + h_x^2 + h_y^2} \, dx \, dy \). \( (4.31) \) is the spacetime version of the same thing.

Here is the short course on spacetime geometry that comes out of Einstein’s special relativity: Just as Euclidean \( \mathbb{R}^3 \) has an ensemble of cartesian coordinate systems \((x, y, z)\), 2 + 1-dimensional spacetime has an ensemble of spacetime coordinate systems \((x, y, t)\) called initial frames. The spatial origin \((x, y) = (0, 0)\) of any inertial frame as seen by another moves at uniform velocity, with speed less than one.

Just as the squared distance \( s^2 := (x' - x)^2 + (y' - y)^2 + (z' - z)^2 \) between two points \((x, y, z), (x', y', z')\) in Euclidean \( \mathbb{R}^3 \) is the same in all cartesian coordinates, so the interval quadratic form

\[ I := (x' - x)^2 + (y' - y)^2 - (t' - t)^2 \]  

(4.32)

between any two events \((x, y, t), (x', y', t')\) in spacetime is the same in all inertial frames. One can well ask: What does this spacetime geometry lesson have to do with our PDE \((4.21)\)? Simply, the wave operator \( \partial_{tt} - \partial_{xx} - \partial_{yy} \) is the same in all inertial frames, just like the Laplacian \( \partial_{xx} + \partial_{yy} + \partial_{zz} \) is the same in all cartesian coordinates. Hence, the non-linear wave equation \((4.21)\) is the same in all inertial frames.

We need to know the physical meaning of the interval quadratic form \((4.31)\). If \( I > 0 \), then there are inertial frames in which events \((x, y, t), (x', y', t')\) are simultaneous \((t = t')\) and \( I \) is the squared spatial distance between them. The displacement between two events with \( I > 0 \) is called spacelike. If \( I < 0 \), there is an inertial frame so that the two events happen at the spatial origin \((x' = x = 0, y' = y = 0)\) and \( \tau := \sqrt{-I} \) is the elapsed time between these events (called the proper time between them). The displacement between two events with \( I < 0 \) is called timelike. In particular, consider uniform motion \( x = ut, y = vt, u^2 + v^2 < 1 \), as seen in some inertial frame. The displacement between \((0, 0, 0)\) and \((ut, vt, t)\) is

---

1Now comes a deep discussion of how “the governance of the world looks the same in all inertial frames, so there is no preferred state of rest”.


timelike and the elapsed proper time is \( \tau = \sqrt{1 - u^2 - v^2} \, t \). This is the time “that passes on a watch of the little guy who moves at uniform velocity from \((x, y) = 0 \) at \( t = 0 \) to \((x, y) = (ut, vt) \) at time \( t > 0 \)”. We are now equipped to explain how the Lagrangian (4.31) is related to “surface area of \( y = h^\partial(x, t) \) in spacetime”.

Figure 4.6 shows the patch of timelike surface \( y = h(x, t) \) which projects to a rectangle of area \( dx \, dt \) in the \( x, t \) plane. Here is how you do the area of this patch: Its base has the usual Euclidean arclength \( ds = \sqrt{1 + h_x^2} \, dx \). Now imagine yourself moving with the front with its normal velocity \( v = \frac{h_t}{\sqrt{1 + h_x^2}} \).

The time elapsed on your pocketwatch as \( t \) increases by \( dt \) is the proper time \( d\tau = \sqrt{1 - v^2} \, dt = \sqrt{1 - \frac{h_t^2}{1 + h_x^2}} \, dt \), and this is the actual spacetime “height” of the patch, and the area = base \( \times \) height is \( \sqrt{1 + h_x^2} - h_t^2 \, dx \, dt \). Apparently the PDE (4.30) is a “minimal surface PDE in spacetime”.

This geometric insight informs the “projection” from the full Lagrangian (4.22) down onto the effective Lagrangian (4.31) of front dynamics. At a formal operational level, here it is: The action associated with a timeslice \( 0 < t < T \) is

\[
S = \int_0^T \int_{\mathbb{R}^2} L \, dy \, dx \, dt.
\] (4.33)

The main contribution to the integral comes from the internal layer about \( \partial D \). In the internal layer we approximate \( L \) by substituting for \( \theta \) the leading order internal layer solution \( \Theta^0 \). Look at the contribution to \( S \) from a patch of \( \partial D \) with \( x_1 < x < x_2, t_1 < t < t_2 \) where \( y = h(x, t) \) is well defined: The
approximation to $L$ in the internal layer is

$$L \sim \frac{1}{2}(h_0^2 - h_t^2 - 1)\Theta_Y^2 - \sin^2 \Theta^0.$$ 

We use (4.29) to substitute for $\sin \Theta^0$, so

$$L \sim -(1 + h_0^2 - h_t^2)\Theta_Y^2$$

and integrating over $Y$ from $-\infty$ to $\infty$ gives

$$\int_{-\infty}^{\infty} LdY \sim -2\sqrt{2} \sqrt{1 + h_0^2 - h_t^2}$$

and the contribution to $S$ from $x_1 < x < x_2$, $t_1 < t < t_2$ is

$$-2\sqrt{2} \int_{t_1}^{t_2} \int_{x_1}^{x_2} \sqrt{1 + h_0^2 - h_t^2} dx dt.$$ 

Of course we can’t represent the whole cylindrical surface $\partial D$ by $y = h^0(x,t)$. But the coordinate-free content is clear: In the limit $\epsilon \to 0$, the action (4.33) is the spacetime surface area of $\partial D$ in the timeslice $0 < t < T$, modulo a multiplicative constant.

**Singularly perturbed eigenvalue problem**

The function $\psi(x)$ defined for $x$ in bounded region $D$ of $\mathbb{R}^3$ and the real number $E$ are an eigenpair of the Laplacian with zero boundary condition,

$$\Delta \psi + E \psi = 0 \quad \text{in} \quad D, \quad \text{(4.34)}$$

$$\psi = 0 \quad \text{on} \quad \partial D. \quad \text{(4.35)}$$

We analyze the ground state eigenpair with the smallest positive $E$. The singular character of the problem comes from the geometry of $D$, illustrated in Figure 4.7. $D$ is a simply connected region minus a ball of radius $\epsilon$ about some interior point $x = q$.

For the first exercise, we take $D$ to be the shell $\epsilon < r := |x| < \pi$. We have a simple exact solution: The spherically symmetric solution of (5.1) which vanishes at $r = \epsilon$ is

$$\psi = \frac{\sin \sqrt{E}(r - \epsilon)}{r}, \quad \text{(4.36)}$$
modulo a multiplicative constant, and the zero boundary condition at \( r = \pi \) implies \( \sqrt{E}(\pi - \epsilon) = n\pi \), where \( n \) is an integer. The ground state eigenvalue has \( n = 1 \), so

\[
E = \left( \frac{\pi}{\pi - \epsilon} \right)^2 = 1 + \frac{2}{\pi} + \frac{\epsilon^2}{3\pi^2} + \ldots .
\] (4.37)

In the outer limit \( \epsilon \to 0 \), \( 0 < r < \pi \) fixed, the eigenfunction (4.36) converges to the ground state eigenfunction of the ball \( r < 1 \) with no hole in the center, and the \( \epsilon \to 0 \) limit of the eigenvalue is 1, also corresponding to the unit ball with no hole. In this sense, the effect of the hole “disappears” as \( \epsilon \to 0 \).

We now reproduce the \( O(\epsilon) \) term in the expansion (4.37) of the eigenvalue by construction and matching of inner and outer expansions. In the spherically symmetric case, (4.34) reduces to

\[
\psi_{rr} + \frac{2}{r} \psi_r + E\psi = 0 .
\] (4.38)

We assume that \( E \) has a power series in \( \epsilon \) whose leading term is the no-hole value 1, so

\[
E = 1 + \epsilon E^1 + \ldots .
\]

The two-term outer expansion of \( \psi \) as \( \epsilon \to 0 \) with \( 0 < r < \pi \) fixed is

\[
\psi \sim \psi^0 + \epsilon\psi^1 ,
\]

where \( \psi^0 =: \frac{\sin \pi r}{r} \) is the no-hole eigenfunction. The perturbation ODE for \( \psi^1 \) is

\[
\psi_{rr}^1 + \frac{2}{r} \psi_r^1 + \psi^1 = -E^1\psi^0 ,
\]
and solutions which vanish at \( r = \pi \) are
\[
\psi^1 = -\frac{E_1}{2r} (\pi - r) \cos r + \alpha \frac{\sin r}{r}
\] (4.39)

where \( \alpha \) is an arbitrary constant. The first term on the right-hand side is a particular solution, and the second is a homogeneous solution which vanishes at \( r = \pi \). Hence, the two-term inner expansion is
\[
\psi \sim \frac{\sin r}{r} + \epsilon \left\{ \alpha \frac{\sin r}{r} - \frac{E_1}{2r} (\pi - r) \cos r \right\}
\] (4.40)

and its inner limit as \( r \to 0 \) is
\[
1 - \frac{\pi E_1}{2} \epsilon + O(r^2 + \epsilon). \tag{4.41}
\]

We anticipate that the \( \frac{1}{r} \) component which diverges as \( r \to 0 \) will be “taken care of” by matching with the inner expansion.

The “inner” representation of \( \psi \) appropriate for \( \epsilon \to 0 \) with \( R := \frac{r}{\epsilon} \) fixed is
\[
\psi = \Psi(R, \epsilon)
\]

and the ODE (4.38) in terms of \( \Psi \) is
\[
\Psi_{RR} + \frac{2}{R} \Psi_R + \epsilon^2 E \Psi = 0
\]
in \( R > 1 \). The one-term inner expansion \( \Psi^0(r) \) satisfies
\[
\Psi^0_{RR} + \frac{2}{R} \Psi^0_R = 0,
\]
subject to \( \Psi^0(1) = 0 \). The solutions are
\[
\Psi^0(R) = A \left( 1 - \frac{1}{R} \right), \tag{4.42}
\]
where \( A \) is an arbitrary constant.

Now, matching: The difference between inner limit (4.40) of two-term outer expansion and on-term inner expansion (4.41) is
\[
1 - \frac{\pi E_1}{2} \epsilon \frac{r}{R} + O(r^2 + \epsilon) - A \left( 1 - \frac{\epsilon}{r} \right) = 1 - A + \frac{\epsilon}{r} \left( A - \frac{\pi E_1}{2} \right) + O(r^2 + \epsilon).
\]
The obvious matching conditions are $A = 1$, $E_1 = \frac{2}{\pi}$. In particular, the $O(\epsilon)$ term of $E$ is $\epsilon^2\frac{2}{\pi}$, which agrees with (4.37). The overlap domain is $r = O(\epsilon^p)$ so $\frac{\epsilon}{r} = O(\epsilon^{1-p}) \gg r^2 + \epsilon = O(\epsilon^{2p} + \epsilon)$, which implies $0 < p < \frac{1}{3}$.

Notice that the constant $\alpha$ in outer solution remains undetermined. Since eigenfunctions are determined only up to a multiplicative constant, we need to impose some additional normalization condition to make it unique. To correctly compute the expansion of eigenvalue, it does not matter what specific normalization you use, so long as it is implemented order-by-order in the perturbation theory. The traditional square integral normalization,

$$\int_\epsilon^\pi \psi^2 4\pi r^2 dr = 1$$

is clearly an unpleasant gentleman. Instead, we use $\psi'(\pi, \epsilon) = \text{given constant independent of } \epsilon$. The outer solution (4.40) is sufficient for the asymptotic evaluation of $\psi'(\pi, \epsilon)$, and we find $\psi'(\pi, \epsilon) \sim -\frac{1}{\pi} + \epsilon^{1-\alpha}$, so $\alpha = 1$.

Next, we break the spherical symmetry of $D$. We take $D = D_0 - B_\epsilon$ where $D_0$ is any bounded, simply connected region, and $B_\epsilon$ is a ball of radius $\epsilon$ about $x = q$ in the interior of $D_0$. $\epsilon$ is so small that $B_\epsilon$ is contained in $D_0$. We assume that the groundstate eigenpair $E^0, \psi^0(x)$ is known for the no-hole case $D = D_0$.

The problem is to determine $E^1$ in the two-term expansion of the eigenvalue, $E \sim E^0 + \epsilon E^1$. Substituting into PDE (4.34) $E \sim E^0 + \epsilon E^1$ and the two-term outer expansion of $\psi$, $\psi \sim \psi^0(x) + \epsilon \psi^1(x)$, we obtain the first-order perturbation equation

$$\Delta \psi^1 + E^0 \psi^1 = -E^1 \psi^0$$

in $D_0 - \{q\}$. We require solutions for $\psi^1$ which vanish on the outer boundary $\partial D_0$,

$$\psi^1 = 0 \quad \text{on} \quad \partial D_0$$

and have $r \psi^1$ bounded as $r = ||x - q|| \to 0$.

A more precise sense of the $x \to q$ behavior of $\psi^1$ is determined from matching with the inner solution. The inner solution is represented by

$$\psi \sim \Psi \left( X = \frac{x - q}{\epsilon}, \epsilon \right)$$

and the PDE (4.34) in terms of $\Psi(X, \epsilon)$ reads

$$\Delta_X \Psi + \epsilon^2 E \Psi = 0$$
where $\Delta_X := \frac{\partial^2}{\partial X_j \partial X_j}$ is Laplacian with respect to $X$. The leading order inner expansion $\Psi^0(X)$ is a harmonic function which vanishes on $R := |X| = 1$, and leading order matching with the outer solution implies $\Psi^0(X) \to \psi^0(q)$ as $R \to \infty$. It is not rocket science to see that

$$\Psi^0 = \psi^0(q) \left(1 - \frac{1}{R} \right) = \psi^0(q) - \epsilon \frac{\psi^0(q)}{r}.$$  \hfill (4.45)

Matching of the two-term outer expansion with this one-term inner expansion gives an effective boundary condition on $\psi^1$ as $r \to 0$,

$$\psi^1 \sim \frac{\psi^0(q)}{r} \quad \text{as} \quad r \to 0. \quad \hfill (4.46)$$

In principle, the program is to determine a particular solution of the linear, inhomogeneous PDE (4.42) for $\psi^1$ subject to the zero boundary condition (4.44) on $\partial D_0$, with $r\psi^1$ bounded as $r =: |x-q| \to 0$. The solution presumably contains $E^1$ as a parameter, and we expect that $E^1$ is determined by enforcing the “matching” boundary condition (4.45). In practice we can carry out this drill explicitly only for a few special cases (such as spherically symmetric $D$). We need a bigger knife.

A piece of linear functional analysis provides: Take the inner product of the PDE (4.43) with $\psi^0$,

$$\langle \psi^0, \Delta \psi^1 + E^0 \psi^1 \rangle := \int_{D_\delta} \psi^0 (\Delta \psi^1 + E^0 \psi^1) \, dx \quad \hfill (4.47)$$

where (and this is the “trick”) $D_\delta := D_0 - B_\delta$. The positive parameter $\delta$ (no relation to $\epsilon$) is so small that $B_\delta$ is in $D_0$. Integrating the vector identity

$$\nabla \cdot (\psi^0 \nabla \psi^1 - \psi^1 \nabla \psi^0) = \psi^0 \Delta \psi^1 - \psi^1 \Delta \psi^0$$

over $D_\delta$ and using the divergence theorem and the zero boundary condition on $\partial D_0$, we find

$$\int_{D_\delta} \psi^0 \Delta \psi^1 \, dx = \int_{D_\delta} \Delta \psi^0 \psi^1 \, dx - \int_{r = |x-q| = \delta} (\psi^0 \psi^1_r - \psi^1 \psi^0_r) \, da.$$
Hence, (4.47) becomes

\[ \int_{D_\delta} (\Delta \psi^0 + E^0 \psi^0) \psi^1 = -E^1 \int_{D_\delta} (\psi^0)^2 dx + \int_{r=\delta} (\psi^0 \psi^1_r - \psi^1 \psi^0_r) da. \]

The left-hand side vanishes because \( \Delta \psi^0 + E^0 \psi^0 = 0 \) in \( D_0 \) (which contains \( D_\delta \), and so

\[ E^1 = \frac{\int_{r=\delta} (\psi^0 \psi^1_r - \psi^1 \psi^0_r) da}{\int_{D_\delta} (\psi^0)^2 dx}. \] (4.48)

As \( \delta \to 0 \), the denominator converges to \( \int_{D_0} (\psi^0)^2 dx \). The “matching” boundary condition (4.45) determines the \( \delta \to 0 \) limit of the surface integral:

We have

\[ \psi^0 \psi^1_r - \psi^1 \psi^0_r = \frac{(\psi^0(q))^2}{r^2} + O\left(\frac{1}{r}\right) \]

so the surface integral in (4.47) converges to \( 4\pi (\psi^0(q))^2 \) as \( \delta \to 0 \). In summary we have

\[ E^1 = \frac{4\pi (\psi^0(q))^2}{\int_{D_0} (\psi^0)^2 dx}. \] (4.49)

**Homogenization of “Swiss Cheese”**

Now suppose \( D \) is a simply connected region minus \( N \gg 1 \) disjoint balls of radius \( \epsilon \), each contained in \( D \). Since the eigenvalue perturbation due to one ball alone is \( O(\epsilon) \), we expect the perturbation due to \( N \) balls in \( O(N\epsilon) \). We aren’t so naive to believe that simple additivity of eigenvalue perturbations is quantitative, but we do expect the order of magnitude is right. Hence we expect that the eigenvalue perturbation is \( O(1) \) if \( N = O\left(\frac{1}{\epsilon}\right) \). We imagine the holes randomly scattered throughout \( D \) with some macroscopic distribution:

Let \( S \) be a closed surface in \( D \) with the topology of a sphere, and \( B^0 \) be its interior. The expected number of holes in \( B^0 \) is

\[ N(B^0) = \frac{1}{\epsilon} \int_{B^0} \rho(x) dx, \]

where \( \rho(x) \) is a macroscopic density function scaled by \( \frac{1}{\epsilon} \).

What is the structure of \( \psi(x, \epsilon) \) in this “Swiss cheese”? The expected interhole distance is \( O\left(\epsilon^{\frac{1}{3}}\right) \). (Each hole is a “cat” who owns a “backyard” of volume \( O(\epsilon) \).) Since \( \epsilon^{\frac{1}{3}} \gg \epsilon \) as \( \epsilon \to 0 \), we expect that in most of the space
between holes, $\psi(x, \epsilon)$ has some leading order outer approximation $\psi^0(x)$, and in an $O(\epsilon)$ neighborhood of each hole, an inner expansion like (4.45).

We analyze the effect of the holes on the eigenvalue $E$. Let $B$ be the portion of $D$ inside the aforementioned surface $S$ in $D$. (The actual piece of Swiss cheese inside $S$.) Integrating the PDE $\Delta \psi + E \psi = 0$ over $B$,

$$\int_{\partial B} \psi_n da + E \int_B \psi dx = 0. \quad (4.50)$$

We carry out a heuristic evaluation of (4.50) as $\epsilon \to 0$. First, due to the “sparseness” of holes, we expect

$$\int_B \psi dx \to \int_{B^0} \psi^0 dx \quad (4.51)$$

where $B^0$ denotes the “no holes” interior of $S$. The piece de resistance is the boundary integral in (4.50). $\partial B$ is the “outer perimeter” component $S$, plus the spheres of radius $\epsilon$ bounding the holes inside $S$, as depicted in Figure 4.8. The surface integral over $S$ presumably converges to $\int_S \psi^0_n da$. The

![Figure 4.8](image)

contribution from a hole centered about $x = q$ is asymptotic to $-4\pi \epsilon \psi^0(q)$. This is obviously based on the inner solution (4.45). In summary, we propose

$$\int_{\partial B} \psi_n da \sim \int_S \psi^0_n da - 4\pi \epsilon \sum_{q \text{ in } B^0} \psi^0(q)$$

as $\epsilon \to 0$. Evoking the macroscopic density $\frac{\delta(x)}{\epsilon}$ of holes we approximate the sum by an integral to obtain

$$\int_{\partial B} \psi_n da \sim \int_S \psi^0_n da - 4\pi \int_{B^0} \rho(x) \psi^0(x) dx.$$
Hence the $\epsilon \to 0$ limit of (4.50) is
\[ \int_S \psi_0^0 da + \int_{B^0} (E^0 - 4\pi \rho(x))\psi_0^0(x) dx = 0, \]
or with the help of the divergence theorem,
\[ \int_{B^0} \{\Delta \psi_0^0 + (E^0 - 4\pi \rho(x))\psi_0^0\} dx = 0. \]

Here, $E^0$ is the leading approximation to the eigenvalue $E$. Since $B^0$ is any subregion of $D^0 := D + \text{holes}$, we have the PDE
\[ \Delta \psi_0^0 + (E^0 - 4\pi \rho(x))\psi_0^0 = 0 \] (4.52)
in $D^0$. The bottom line: we have an effective eigenvalue problem of $\psi_0^0(x), E^0$ consisting of the PDE (4.52) subject to the zero boundary condition $\psi_0^0 = 0$ on $\partial D^0$. 
Chapter 5

Prant’l boundary layer theory

We begin by formulating the equations for steady two-dimensional flow about a streamlined body $B$ as depicted in Figure 5.1. “Two dimensional” means that all state variables depend only on $x_1$ and $x_2$ and in particular, the velocity field is

$$ u(x := x_1 e_1 + x_2 e_2) = u_1(x)e_1 + u_2(x)e_2 $$

(5.1)

Figure 5.1

with no $e_3$ component. The “incoming” velocity field at $x_1 = -\infty$ is $Ue_1$, with $U > 0$. The body is symmetric about the $x_1$ axis so it is sufficient to describe the portion of flow in the upper half plane.

We assume uniform density, incompressible fluid. Incompressible means that the area of a moving region consisting of the same fluid particles is
constant, independent of time. Let \( D(t) \) denote any such moving region at time \( t \), depicted in Figure 5.1. The area of \( D(t) \) is the line integral

\[
A = - \int_{\partial D(t)} x_2 dx_1,
\]

where \( \partial D(t) \) has counterclockwise orientation. Let \( \mathbf{x}(s,t) \) be a parametric representation of the boundary curve \( \partial D(t) \), with \( s, 0 < s < 1 \), labeling fluid particles on the boundary. As \( s \) ranges between zero and one, we cover the closed curve \( \partial D(t) \) once. In particular \( \mathbf{x}(0,t) = \mathbf{x}(1,t) \). Since \( \mathbf{x}(s,t) \) with \( s \) fixed represents the position of a fluid particle as a function of time, and fluid particles are carried with velocity \( \mathbf{u}(\mathbf{x}) \), we have

\[
\partial_t \mathbf{x}(s,t) = \mathbf{u}(\mathbf{x}(s,t)).
\]

The parametric representation of the line integral (5.2) is

\[
A = - \int_0^1 x_2(s,t) \partial_s x_1(s,t) ds.
\]

Since the area is independent of time, time differentiation of (5.4) gives

\[
0 = \int_0^1 \{ \partial_t x_2 \partial_s x_1 + x_2 \partial_s x_1 \} ds,
\]

and integration by parts applied to the second term gives

\[
0 = \int_0^1 \{ \partial_t x_2 \partial_s x_1 - \partial_t x_1 \partial_s x_2 \} ds.
\]

We have \( \partial_t x_1 = u_1(\mathbf{x}), \partial_t x_2 = u_2(\mathbf{x}) \) by (5.3), so

\[
0 = \int_0^1 \{ u_2 \partial_s x_1 - u_1 \partial_s x_2 \} ds = \int_{\partial D} u_2 d x_1 - u_1 d x_2.
\]

By Green’s theorem the right-hand side is \( -\int_D (\partial_1 u_1 + \partial_2 u_2) d\mathbf{u} \). Since this integral vanishes for all moving regions \( D \) of fluid, we have

\[
\nabla \cdot \mathbf{u} := \partial_1 u_1 + \partial_2 u_2 = 0
\]

at all points in the fluid.
Chapter 5. Prantl boundary layer theory

The influences of physical forces such as pressure and viscous stress upon the flow are quantified by “Newton’s second law expressed locally”:

“Acceleration of a fluid particle = force per unit mass”. \hspace{1cm} (5.6)

First, we explain how to compute accelerations of fluid particles in terms of velocity field $\mathbf{u}(\mathbf{x})$. Let $\mathbf{x}(t)$ be the position of a fluid particle as a function of time. Since fluid particles are “carried in the flow $\mathbf{u}(\mathbf{x})$”,

$$\dot{\mathbf{x}}(t) = \mathbf{u}(\mathbf{x}(t)).$$ \hspace{1cm} (5.7)

Time differentiation of the $i$-th component of (5.7), use of the chain rule and (5.7) gives

$$\ddot{x}_i(t) = (\partial_j u_i)(\mathbf{x}(t))\dot{x}_j(t) = (u_j \partial_j u_i)(\mathbf{x}(t)), \hspace{1cm} ^1$$

or in vector notation

$$\ddot{\mathbf{x}}(t) = ((\mathbf{u} \cdot \nabla)\mathbf{u})(\mathbf{x}(t)).$$ \hspace{1cm} (5.8)

(5.8) quantifies the left-hand side of the “local Newton’s law” (5.6).

The “force per unit mass” in the right-hand side of (5.6) has two components, from pressure and from viscosity. Let $p(\mathbf{x})$ be the pressure field. The pressure force acting on the fluid in region $D$ in $\mathbb{R}^3$ is

$$-\int_{\partial D} p \mathbf{n} \, da.$$ \hspace{1cm} (5.9)

Figure 5.2 is the “visual explanation”. The $i$-th component is

\footnotesize
\[ ^1 \text{Sum over repeated index } j. \]
Chapter 5. Prantl boundary layer theory

Figure 5.2

\[ -\int_{\partial D} p e_i \cdot n da \]

and by the divergence theorem it is

\[ -\int_D \nabla \cdot (p e_i) dx = -\int_D \partial_i p dx. \]

In vector notation

\[ -\int_{\partial D} p n da = -\int_D \nabla p dx. \]

We see that \(-\nabla p\) is pressure force per unit volume, and dividing by the density \(\rho\), we get \(-\frac{1}{\rho} \nabla p\) as the pressure force per unit mass.

The viscous stress (force per unit area) acting on \(D\) across the “patch” in Figure 5.2 is \(T(n da)\) where \(T\) is the linear transformation that converts the area element \(n da\) into force per unit area. Fluid people call \(T\) the *viscous stress tensor*. Viscous stresses are generated by *gradients* in the velocity field. In cartesian coordinates, the matrix representing \(T\) in incompressible flow is

\[ T_{ij} = \eta (\partial_i u_j + \partial_j u_i). \]  

(5.10)

Here, the positive constant \(\eta\) is called the *dynamic viscosity*. The viscous force acting on \(D\) is

\[ \int_{\partial D} T n da, \]
Chapter 5. Prandtl boundary layer theory

with \( i \)-th component

\[
\int_{\partial D} (Tn)_i da = \int_{\partial D} T_{ij}n_j da
= \int_{\partial D} (T_{ij}e_j) \cdot nda
= \int_D \nabla \cdot (T_{ij}e_j) dx
= \int_D \nabla \nabla \cdot (T_{ij}e_j) dx
= \int_D \partial_j T_{ij} dx.
\] (5.11)

The third equality uses the divergence theorem. Substituting (5.10) for \( T_{ij} \) into the right-hand side of (5.11), we find

\[
\int_{\partial D} (Tn)_i da = \eta \int_D \{ \partial_i (\partial_j u_j) + \partial_{jj} u_i \} dx
= \eta \int_D \Delta u_i dx.
\]

The second equality uses \( \partial_j u_j = \nabla \cdot u = 0 \), and \( \Delta := \partial_{jj} \) is the Laplacian. In summary,

\[
\int_{\partial D} Tn da = \eta \int_D \Delta u dx,
\]

so \( \eta \Delta u \) is the viscous force per unit volume. The viscous force per unit mass is \( \nu \Delta u \), where \( \nu := \frac{\eta}{\rho} \) is called the kinematic viscosity.

The total force per unit mass is

\[
-\frac{1}{\rho} \nabla p + \nu \Delta u
\] (5.12)

the sum of pressure and viscous components. Equating the acceleration in (5.8) to the force per unit mass in (5.12), we obtain the (steady) Navier–Stokes equation,

\[
(u \cdot \nabla) u = -\frac{1}{\rho} \nabla p + \nu \Delta u
\] (5.13)

as the quantitative expression of the “local Newton’s law” (5.6).

There are boundary conditions on the velocity field \( u \). Along the surface of the body \( B \), there is the no-slip condition

\[
u = 0 \quad \text{on} \quad \partial B.
\] (5.14)
The symmetry of the flow about the $x_1$ axis dictates

$$u_2 = 0 \quad (5.15)$$

on the portion of $x_1$ axis not occupied by the body. Finally, we assume uniform flow in the $e_1$ direction far from the body, so

$$\mathbf{u} \to U e_1 \quad \text{as} \quad r =: \sqrt{x_1^2 + x_2^2} \to \infty. \quad (5.16)$$

Here, $U$ is a positive constant. Notice that there are no explicit boundary conditions on the pressure $p$. Nevertheless, the PDE (5.5), (5.13) and boundary conditions (5.14)–(5.16) constitute a boundary value problem that determines $\mathbf{u}$ uniquely, and $p$ up to a constant.

**Stream function and vorticity**

The “streamlines” in Figure 5.1 are integral curves of the velocity field. Consider the region $R$ corresponding to the “river” of fluid bounded above by a given streamline, and below by the $x_1$ axis, or the body, as depicted in Figure 5.3. Take $C$ to be any curve that crosses the “river” from “bottom” to

![Figure 5.3](image)

“top”. Four possibilities are depicted in Figure 5.3. The area of fluid which sweeps past $C$ per unit time is the line integral

$$\int_C \mathbf{u} \cdot \mathbf{n} ds = \int_C (u_1 e_1 + u_2 e_2) \cdot (dx_2 e_1 - dx_1 e_2)$$

$$= \int_C -u_2 dx_1 + u_1 dx_2. \quad (5.17)$$

Due to incompressibility, the flux (5.17) is independent of $C$, so each streamline is parametrized by the common value of all these flux line integrals. In this sense, the streamlines are the level curves of a *stream function* $\psi(\mathbf{x})$.
which is the value of the line integral (5.17) along any curve \( C \) from the “bottom” (heavy curve in Figure 5.3) to \( \mathbf{x} \). It follows from (5.17) that the velocity field components are derivatives of \( \psi \),

\[
    u_1 = \partial_2 \psi, \quad u_2 = -\partial_1 \psi. \tag{5.18}
\]

The stream function representation (5.18) of \( \mathbf{u} \) “eliminates” the incompressibility equation \( \nabla \cdot \mathbf{u} = 0 \). For any sufficiently smooth \( \psi(\mathbf{x}) \) we have

\[
    \nabla \cdot \mathbf{u} = \partial_1 u_1 + \partial_2 u_2 = \partial_{12} \psi - \partial_{21} \psi = 0.
\]

It remains to derive from the Navier–Stokes equation (5.13) a PDE for \( \psi(\mathbf{x}) \).\(^2\) This is best understood in terms of the vorticity

\[
    \omega := \partial_1 u_2 - \partial_2 u_1. \tag{5.19}
\]

Here is its physical meaning: In the frame of reference of a particle moving with the fluid, the relative velocity field asymptotes to its linearization as \( \mathbf{x} \to 0 \), \( \mathbf{u} \sim \mathbf{Jx} \), where \( \mathbf{J} \) is the Jacobian of \( \mathbf{u} \) at \( \mathbf{x} = 0 \), with components \( J_{ij} = (\partial_i u_j)(0) \). The antisymmetric part of \( \mathbf{J} \) is \( \begin{bmatrix} 0 & -\omega_2 \\ \omega_1 & 0 \end{bmatrix} \), which represents the velocity field of solid body rotation at angular velocity \( \omega_2 \). Hence, vorticity measures the local rotation component of the flow. In incompressible flow, vorticity is a “source” of the stream function in the sense that

\[
    \Delta \psi = \partial_1 \partial_1 \psi + \partial_2 \partial_2 \psi \\
    = -\partial_1 u_2 + \partial_2 u_1 = -\omega. \tag{5.20}
\]

The Navier–Stokes equation (5.13) implies a nice transport equation for the vorticity. By taking the \( x_1 \) derivative of the \( x_2 \)-component of Navier–Stokes, minus the \( x_2 \) derivative of \( x_1 \)-component, the pressure \( p \) is eliminated, and what’s left is

\[
    (\mathbf{u} \cdot \nabla) \omega = \nu \Delta \omega. \tag{5.21}
\]

In transport terminology, we say that “vorticity is convected with the flow \( \mathbf{u} \), and diffuses with diffusion coefficient \( \nu \)”.

The stage has been finally set to derive a well-posed boundary value problem for \( \psi(\mathbf{x}) \). First, set \( \omega = -\Delta \psi \) in the vorticity transport equation (5.21), to obtain

\[
    (\mathbf{u} \cdot \nabla) \Delta \psi = \nu \Delta^2 \psi. \tag{5.22}
\]

\(^2\)What happens to \( p \)? We’ll get rid of it. You’ll see.
Chapter 5. Prant'l boundary layer theory

Here, the components of $\mathbf{u}$ are given in terms of $\psi$ according to (5.18) so (5.22) is a single (nasty) PDE for $\psi$. Now the boundary conditions: By definition of $\psi$,

$$\psi \equiv 0 \quad (5.23)$$
on the “bottom” consisting of $x_1$ axis, or “skin” of the body. On the skin of the body, $\mathbf{u}$ is zero, so both first derivatives, and hence the normal derivative vanish there:

$$\partial_n \psi = 0 \quad (5.24)$$
on the “skin” $\partial B$. Finally, the uniform flow at infinity is represented by the effective boundary conditions

$$\partial_2 \psi = u_1 \to U, \quad -\partial_1 \psi = u_2 \to 0 \quad (5.25)$$
as $r := \sqrt{x_1^2 + x_2^2} \to \infty$. In summary, the Navier–Stokes boundary value problem for $\mathbf{u}$, $p$ has been reduced to a boundary value problem for the stream function $\psi(\mathbf{x})$. There is only one state variable, but the PDE is non-linear and fourth-order.

Preliminary non-dimensionalization

The “standard” units of the variables are $[x]$ (scaling unit of $\mathbf{x}$) = $L$ (length of body), and $[\psi] = UL$. Recall that $\psi$ as “rate of area sweeping by” implies that its units are $(\text{length})^2 / \text{time}$, and $LU$ is the obvious candidate. The non-dimensional form of the PDE (5.22) is

$$(\mathbf{u} \cdot \nabla) \Delta \psi = \epsilon^2 \Delta^2 \psi. \quad (5.26)$$

Here, the dimensionless velocity $\mathbf{u}$ is given in terms of dimensionless $\psi$ by (5.18) and $\epsilon^2$ is the inverse of the famous Reynold’s number,

$$\epsilon^2 := \frac{1}{Re} = \frac{\nu}{LU}. \quad (5.27)$$
The “bottom” and “skin” boundary conditions (5.23), (5.24) are invariant under scaling so they stand as they are. The effective boundary conditions at infinity become

$$\partial_2 \psi \to 1, \quad \partial_1 \psi \to 0 \quad (5.28)$$
as $r \to \infty$.

As discussed in the introduction to Chapter 3, the usual limit for large, rapidly moving bodies is high Reynold’s number, so we consider
\( \epsilon := \frac{1}{\sqrt{Re}} \to 0 \). Recall that the thickness of boundary layer adjacent to the surface of the body is on the order of \( \frac{L}{\sqrt{Re}} = \epsilon L \), so the dimensionless boundary layer thickness is \( O(\epsilon) \). We construct matched asymptotic expansions of \( \psi \): An outer expansion \( (\epsilon \to 0 \text{ with } x \text{ fixed}) \) and an inner expansion \( (\epsilon \to 0 \text{ with normal displacement from surface } = O(\epsilon)) \).

**Outer expansion and “dry water”**

We examine a two-term outer expansion of \( \psi \),

\[
\psi \sim \psi^0(x) + \epsilon \psi^1(x). \tag{5.29}
\]

The leading order term \( \psi^0(x) \) is presumably the stream function of the inviscid (\( \epsilon = 0 \)) flow about the body. (In the language of Dr. Feynman, inviscid fluid is called “dry water”.) The PDE (5.26) by itself suggests an asymptotic expansion of \( \psi \) with only even powers of \( \epsilon \), but due to the \( O(\epsilon) \) boundary layer thickness, we’ll see that asymptotic matching between outer and inner expansions induces an \( O(\epsilon) \) component of outer expansion. There is a physical idea here: The effectively inviscid flow away from the body “sees” the boundary layer as an additional thickness of body, and the two-term expansion \( \psi^0 + \epsilon \psi^1 \) is the stream function of an inviscid flow about this “fattened” body.

We formulate the reduced boundary value problem for the inviscid (\( \epsilon = 0 \)) flow. The obvious \( \epsilon = 0 \) reduction of (5.26) is

\[
(u^0 \cdot \nabla)\Delta \psi^0 = 0, \tag{5.30}
\]

where the inviscid velocity field \( u^0 \) is related to \( \psi^0 \) according to (5.18). We see that the leading order vorticity \( -\Delta \psi^0 \) is constant along the streamlines of \( u^0 \). These streamlines extend to \( x_1 = -\infty \) where there is no vorticity, so the leading order flow \( u^0 \) is irrotational (zero vorticity) and hence, \( \psi^0 \) harmonic,

\[
\Delta \psi^0 = 0.
\]

\( \psi^0 \) satisfies the boundary conditions (5.28) at infinity. The inviscid flow is expected to be symmetric about the \( x_1 \) axis, like in Figure 5.1, and this is automatic if \( \psi^0 \) is odd in \( x_1 \). Hence, we have \( \psi^0 = 0 \) along the portion of \( x_1 \) axis not occupied by body. The inviscid flow does not cross the “skin” of the body, so the velocity \( u^0 \) should be tangential along the skin. This means that the \( \psi^0 = 0 \) streamline continues in \( 0 < x_1 < 1 \) by following the skin. Hence,
\( \psi^0 \) vanishes along the whole “bottom” curve in Figure 5.3. We now have a complete boundary value problem for \( \psi^0 \) that can be solved in principle, and the streamlines that derive from \( \psi^0 \) have a qualitative resemblance to the streamlines of the viscous flow.

The perturbation equation for \( \psi^1(x) \) is

\[
(u^0 \cdot \nabla) \Delta \psi^1 = -(u^1 \cdot \nabla) \Delta \psi^0.
\]

We already know \( \Delta \psi^0 \equiv 0 \), so the directional derivative of \( \Delta \psi^1 \) along the leading order streamlines is zero, and repeating the argument for \( \Delta \psi^0 \equiv 0 \), we find \( \Delta \psi^1 \equiv 0 \), so \( \psi^1 \) is harmonic as well. \( \psi^1 \) satisfies homogeneous boundary conditions at infinity, \( \partial_1 \psi, \partial_2 \psi \to 0 \) as \( r \to \infty \). An effective boundary condition on \( \psi^1 \) along the body results from asymptotic matching of the two-term outer expansion (5.29) with the one-term inner expansion.

As mentioned before, \( \psi^1 \) can be regarded as the perturbation of the inviscid flow induced by a slight “fattening” of the body.

**Inner expansion**

First, we determine the scaling of \( \psi \) in the boundary layer. The leading order outer expansion \( \psi^0 \) is zero on the skin, with linear contact: The tangential velocity of inviscid flow along the skin is the normal derivative of \( \psi^0 \), generally non-zero. Hence, the leading order outer solution \( \psi^0 \) is \( O(\epsilon) \) in the boundary layer of \( O(\epsilon) \) thickness. This suggests that the inner solution in the boundary layer is \( O(\epsilon) \), and we introduce \( \Psi \) as the stream function measured in units of \( \epsilon \), so

\[
\psi = \epsilon \Psi. \tag{5.31}
\]

The spatial coordinates of the inner expansion are \( s := \text{arclength along skin, measured from the leading edge, and } y, \text{ the normal displacement from skin measured in units of } \epsilon \). The geometry of \( s, y \) coordinates is depicted in Figure 5.4. Let \( X(s) \) be the parametric representation of the upper boundary curve with respect to arclength \( s \), and define unit vectors

\[
\hat{s} := X_s(s) = (\partial_s X_1, \partial_s X_2), \quad \hat{y} = (-\partial_s X_2, \partial_s X_1)
\]

(analogous to \( r \) and \( \theta \) for polar coordinates). Spatial positions \( x \) are represented in terms of \( s \) and \( y \) by

\[
x = X(s) + \epsilon y \hat{y}(s). \tag{5.32}
\]
The agenda is to convert (5.26) into a PDE for $\Psi = \Psi(s, y)$.

Here are the background calculations: The gradient operator in $s, y$ coordinates is

$$
\nabla = \frac{\partial_s}{1 - \epsilon \kappa y} \hat{s} + \frac{1}{\epsilon} \partial_y \hat{y}.
$$

(5.33)

$\kappa = \kappa(s)$ in (5.33) is the curvature of the upper boundary so that $\hat{s}_s = \kappa \hat{y}$. Let $u, v$ be $\hat{s}$ and $\hat{y}$ components of the velocity field $u$. The conversion of (5.18) into $s, y$ coordinates is

$$
u = \hat{y} \cdot \nabla \psi = \frac{1}{\epsilon} \partial_y \psi = \partial_y \Psi,
$$

$$
v = -\hat{s} \cdot \nabla \psi = -\frac{1}{1 - \epsilon \kappa y} \partial_s \psi = -\frac{\epsilon}{1 - \epsilon \kappa y} \partial_s \Psi.
$$

(5.34)

From (5.33), (5.34) we find that the directional derivative operator $u \cdot \nabla$ in (5.20) can be expressed as

$$
u \cdot \nabla = \frac{1}{1 - \epsilon \kappa y} (\partial_y \Psi \partial_s - \partial_s \Psi \partial_y).
$$

(5.35)

An important point: Since $u$ in the boundary layer is mainly in the $s$ direction, one might naively expect that $u \cdot \nabla$ is nearly proportional to $\partial_s$. Although $v$ in (5.34) is smaller than $u$ by a factor of $\epsilon$, remember that $y$ is...
normal displacement in units of $\epsilon$, so derivative in $\hat{y}$ direction is $\frac{1}{\epsilon} \partial_y$, and hence the $\partial_y$ component of $\mathbf{u} \cdot \nabla$ balances the $\partial_s$ component in (5.35).

The Laplacian of a scalar field $h(s, y)$ in $s, y$ coordinates can be calculated from the variational derivative of $\frac{1}{2} \int |\nabla h|^2 dx$. The result is

$$\Delta h = \frac{1}{\epsilon^2} \frac{1}{1 - \epsilon \kappa y} \partial_y((1 - \epsilon \kappa y) \partial_y h) + \frac{1}{(1 - \epsilon \kappa y)^2} \partial_{ss} h. \quad (5.36)$$

The details behind (5.33), (5.36) are served up as an obnoxious exercise. The rationale is that if the professor has to suffer through it, so do you. In summary, we have expressed $\mathbf{u} \cdot \nabla$ and $\Delta \Psi$ in terms of $s$ and $y$ derivatives of $\Psi$, so in principle we have converted (5.20) into a PDE for $\Psi(s, y)$.

At first glance, what a mess! But if we are after the leading order inner expansion only, the mess goes away: In the limit $\epsilon \to 0$, (5.35), (5.36) reduce to

$$\mathbf{u} \cdot \nabla = \partial_y \Psi \partial_s - \partial_s \Psi \partial_y + O(\epsilon),$$

$$\Delta = \frac{1}{\epsilon^2} \partial_{yy} + O\left(\frac{1}{\epsilon}\right),$$

and the reduced PDE for the leading order inner solution $\Psi^0(s, y)$ is

$$(\partial_y \Psi^0 \partial_s - \partial_s \Psi^0 \partial_y) \left(\frac{1}{\epsilon^2} \partial_{yy} \Psi^0\right) = \epsilon^2 \left(\frac{1}{\epsilon^2} \partial_{yy} \Psi^0\right)^2 \Psi^0$$

or

$$(\partial_y \Psi^0 \partial_s - \partial_s \Psi^0 \partial_y) \partial_{yy} \Psi^0 = \partial_{yyy} \Psi^0. \quad (5.37)$$

The no slip condition along the skin is represented by boundary conditions

$$\Psi^0(s, 0) = 0, \quad \partial_y \Psi^0(s, 0) = 0. \quad (5.38)$$

Leading order matching and a “first integral”

An effective boundary condition on $\Psi^0$ as $y \to \infty$ comes from leading order matching between inner and outer approximations to the $\hat{s}$ component of velocity field: The inner approximation $\partial_y \Psi^0(s, y)$ of $\hat{s}$-velocity should asymptote to the tangential velocity

$$U(s) := \partial_n \Psi^0 \quad (5.39)$$

of the inviscid flow along the skin as $y \to \infty$, so we have

$$\partial_y \Psi^0(s, y) \to U(s) \quad \text{as} \quad y \to \infty. \quad (5.40)$$
In addition we expect that higher $y$-derivatives of $\Psi^0$ vanish as $y \to \infty$. In particular, the leading order vorticity $-\frac{1}{\epsilon} \partial_y \Psi^0$ should vanish as $y \to \infty$. With the help of the effective boundary conditions on $\Psi^0$ as $y \to \infty$, we derive a “first integral” of the PDE (5.37) in an exercise. The result is

$$
(\partial_y \Psi^0 \partial_s - \partial_s \Psi^0 \partial_y) \partial_y \Psi^0 - U(s)U'(s) = \Psi^0_{yyy}. \tag{5.41}
$$

In summary, the essential equations of boundary layer theory are the “first integral” PDE (5.41) and the three boundary conditions (5.38), (5.40). The boundary layer equations prescribe the “downstream evolution” of $\Psi^0(s,y)$: Given $\Psi^0$ as a function of $y$ on $s = a$, the boundary layer equations determine $\Psi^0$ in $s > a$. Some physical and mathematical features of this “downstream evolution” are spelled out in an exercise.

**Displacement thickness**

A more refined matching leads to the classical notion of displacement thickness. The outer limit of the leading order inner expansion of $\psi$ is

$$
\epsilon \Psi^0 = \epsilon \int_0^y \Psi^0_y(s,y)\,dy
= \epsilon U(s)y - \epsilon \int_0^y (U(s) - \Psi^0_y(s,y))\,dy
\sim \epsilon U(s)y - \epsilon \int_0^\infty (U(s) - \partial_y \Psi^0(s,y))\,dy \tag{5.42}
$$

as $y \to \infty$. Since $\partial_y \Psi^0$ generally increases from value 0 on $y = 0$ to value $U(s)$ as $y \to \infty$, the integral $\int_0^\infty (U(s) - \Psi^0_y)\,dy$ is positive. The two-term outer expansion expressed as a function of $s$, $y$ is

$$
(\psi^0 + \epsilon \psi^1)(x = X(s) + \epsilon y\hat{y}(s))
$$

and its inner ($y \to 0$) limit is

$$
\epsilon U(s)y + \epsilon \psi^1(s,0) + O(\epsilon^2 y^2). \tag{5.43}
$$

Comparing (5.42) and (5.43) we discern the matching condition

$$
\psi^1(s,0) = -\int_0^\infty (U(s) - \partial_y \Psi^0)\,dy. \tag{5.44}
$$
Chapter 5. Prantl boundary layer theory

It follows from (5.43), (5.44) that the two-term outer expansion of $\psi$ vanishes along a curve above the body, whose approximation in $s, y$ coordinates is

$$y = \int_0^\infty \left( 1 - \frac{\partial_y \Psi^0}{U(s)} \right) \, dy.$$ 

In this sense, the “outer (effectively) inviscid flow feels the boundary layer as an added thickness of body”. Recall that $y$ is normal displacement in units of $\epsilon$, so the normal displacement of the $\psi^0 + \epsilon \psi^1 = 0$ level curve above the body is approximated by

$$\delta := \epsilon \int_0^\infty \left( 1 - \frac{\partial_y \Psi^0}{U(s)} \right) \, dy. \quad (5.45)$$

$\delta$ is called the displacement thickness induced by the boundary layer. Figure 5.5 is a visualization of the displacement thickness.

![Figure 5.5](image)

**Figure 5.5**

**Solutions based on scaling symmetry**

The difficult PDE boundary value problem for $\Psi^0$ reduces to ODE boundary value problems for simple body geometries with scale invariance. In Figure 5.6a, the body is the wedge with azimuthal angles $\theta$ with $|\theta| < \alpha$, where $0 \leq \alpha < \pi$. We’ve depicted streamlines of an inviscid flow “incoming” from $x_1 = -\infty$. This inviscid stream function is harmonic on $\alpha < |\theta| < \pi$ and vanishes on $\theta = \alpha$, and given the symmetry of streamlines about the $x_1$ axis, vanishes on $|\theta| = \pi$ as well. It takes separation of variables from

$$\psi(r, \theta) = \frac{\pi - \alpha}{\pi} U \left( \frac{r}{L} \right)^{\pi - \alpha} \sin \frac{\pi}{\pi - \alpha} \left( \frac{\theta - \alpha}{\pi - \alpha} \right). \quad (5.46)$$
The prefactor $\frac{\pi - \alpha}{\pi}$ is a piece of disingenuous “hindforesight” so it doesn’t appear in later equations as a tedious nuisance. $L$ is a length, and $U$ is a velocity. Using $L$ and $LU$ as units of $r$ and $\psi$, the dimensionless version of (5.46) has $U$ and $L$ replaced by unity. The dimensionless tangential velocity on the skin $|\theta| = \alpha$ as a function of $s := r$ is

$$U(s) = \frac{1}{s} \psi(s, \alpha) = s^m, \quad m := \frac{\alpha}{\pi - \alpha}. \tag{5.47}$$

The inviscid flow does not asymptote to a uniform $x_1$ velocity as $x_1 \rightarrow -\infty$. In fact the flow speed along the negative $x_1$ axis is (5.47) with $s = -x_1$. But the “wedge” flow is realized asymptotically (modulo a multiplicative constant) near the wedgelike leading edge of a finite body, such as depicted in Figure 5.1. There are two compelling special cases: $\alpha = 0, m = 0$ (“flat plate” in Figure 5.6b) and $\alpha = \frac{\pi}{2}, m = 1$ (“wall” in Figure 5.6c).

For any $\alpha, 0 < \alpha < \pi$, the inviscid wedge flow accelerates as we move away from the leading edge. A negative $\alpha$ in (5.47) formally corresponds to a decelerating inviscid flow, but a negative wedge angle can be realized only by physically absurd “surgery” of $\mathbb{R}^2$ which consists of “giving $\mathbb{R}^2$ a wedgie”.\footnote{Freud said “The Irish are the only people who can’t be helped by psychoanalysis.” Surely he didn’t tangle with mathematicians.}
Nevertheless, the case $\alpha < 0$ has been studied as a formal model of boundary layers subject to decelerating outer flow.

With $U(s)$ as in (5.47), the PDE (5.41) for $\Psi^0$ reads

$$\left( \partial_y \Psi_0 \partial_s - \partial_s \Psi_0 \partial_y \right) \partial_y \Psi_0 - ms^{2m-1} = \partial_{yyy} \Psi_0. \quad (5.48)$$

The boundary value problem consisting of PDE (5.48), subject to the boundary conditions (5.38), (5.40), has a scaling symmetry: Suppose we use $a_1$, $a_2$, and $b$ as scaling units of $s$, $y$ and $\Psi^0$ (when $s = a_1$, the new scaled $s$ is 1, similarly for $y$ and $\Psi^0$). Let $\Psi^0(s, y)$ represent the “new” scaled stream function in terms of scaled $s, y$. The “original” stream function in terms of the scaled one is $b\Psi^0 \left( \frac{s}{a_1}, \frac{y}{a_2} \right)$. Given these units of $s, y, \Psi^0$, the convection term $(\partial_y \Psi^0 \partial_s - \partial_s \Psi^0 \partial_y) \partial_y \Psi^0$ in (5.48) gets scaled by factor $\frac{b^2}{a_1 a_2^2}$, $ms^{2m-1}$ gets scaled by $a_1^{2m-1}$, and $\partial_{yyy} \Psi^0$ gets scaled by $\frac{b}{a_2^3}$. The rescaled PDE is the same as the original provided that

$$\frac{b^2}{a_1 a_2^2} = a_1^{2m-1} = \frac{b}{a_2^3}.$$
from which it follows that

\[ a_2 = a \frac{1-m}{2}, \quad b = a_1 \frac{1+m}{2}. \]

The rescaled stream function satisfies the same boundary conditions on the skin \( \Psi^0(s,0), \partial_y \Psi^0(s,0) = 0 \), as the original. The boundary condition \( \partial_y \Psi^0 \to U(s) = s^m \) is also scale invariant: \( \partial_y \Psi^0 \) scales by \( \frac{b}{a_2} = a_1^m \), same as \( s^m \). In summary, if \( \Psi^0(s,y) \) satisfies the boundary layer equations, then so does

\[ b \Psi^0 \left( \frac{s}{a_1}, \frac{y}{a_2} \right) = a_1 \frac{1+m}{2} \Psi^0 \left( \frac{s}{a_1}, \frac{y}{\frac{1-m}{2}} \right), \]

for any \( a_1 > 0 \). If the solution is assumed to be unique, then for all \( s, y \) and \( a_1 > 0 \),

\[ \Psi^0(s,y) = a_1 \frac{1+m}{2} \Psi^0 \left( \frac{s}{a_1}, \frac{y}{\frac{1-m}{2}} \right). \]

In particular, set \( a_1 = s \), so

\[ \Psi^0(s,y) = s \frac{1+m}{2} F \left( \zeta := \frac{y}{\sqrt{s}} \right), \quad (5.49) \]

where \( F(\zeta) := \Psi^0(1,\zeta) \). (5.49) is an example of a similarity solution. Substituting the representation (5.49) into the PDE boundary value problem for \( \Psi^0 \) gives an ODE boundary value problem for the “profile function” \( F(\zeta) \):

\[ m(F')^2 - \frac{m+1}{2} FF'' - m = F''' \quad \text{in} \quad g > 0, \]

\[ F(0) = 0, \quad F'(0) = 0, \quad F'(\infty) = 1. \quad (5.50) \]

In general, solutions are determined numerically.

**Blasius flow over flat plate**

For \( m = 0 \), the similarity solution (5.49) becomes

\[ \Psi^0(s,y) = \sqrt{s} F \left( \zeta := \frac{y}{\sqrt{s}} \right). \]

The leading approximation to the dimensionless stream function in the boundary layer is

\[ \psi \sim \epsilon \Psi^0(s,y) = \epsilon \sqrt{s} F \left( \frac{y}{\sqrt{s}} \right). \]
We convert to dimensional form:

\[
\frac{\psi}{LU} \sim \sqrt{\frac{\nu}{LU}} \sqrt{\frac{s}{L}} F\left(\frac{y / \left(\sqrt{\frac{\nu}{LU}} L\right)}{\sqrt{s / L}}\right)
\]

or

\[
\psi \sim \sqrt{\nu} U s F\left(\frac{U}{\nu s} y\right).
\] (5.51)

We did this conversion, recalling that \(\epsilon := \sqrt{\frac{\nu}{LU}}\), that the units of \(s, y\) and \(\psi\) are \(L, \epsilon L\) and \(LU\), where \(L\) is a length and \(U\) a velocity. Notice that in (5.51) \(L\) has dropped out. Why? When we replace the finite length body with a semi-infinite flat plate, we remove the only length parameter from the problem. In the language of physics, “the flow has no intrinsic length scale”. The only parameters we have are \(\nu\) and \(U\), and only those appear in (5.51). The leading approximation to the velocity field in the boundary layer based on (5.50) is the famous *Blasius flow*

\[
u = \partial_y \psi \hat{y} \sim U F'\left(\frac{U}{\nu s} y\right) \hat{y}.
\] (5.52)

Figure 5.8 is Batchelor’s graph of the velocity profile \(F'(\zeta)\), based on the numerical solution of the boundary value problem (5.50) for \(m = 0\). The dimensionless displacement thickness of the Blasius boundary layer based on
(5.45) is
\[ s = \epsilon \int_0^{\infty} \left( 1 - F' \left( \frac{y}{\sqrt{s}} \right) \right) dy \]
\[ = \epsilon \sqrt{s} \int_0^{\infty} (1 - F' (\zeta)) d\zeta \simeq 1.72 \epsilon \sqrt{s}, \]
and the dimensional displacement thickness, determined from
\[ \frac{s}{L} \simeq 1.72 \sqrt{\frac{\nu}{LU}} \sqrt{\frac{s}{L}}, \]
is
\[ s \simeq 1.72 \sqrt{\frac{\nu s}{U}}. \]

\( m \neq 0 \)

Figure 5.9 shows velocity profiles based on numerical solutions to the boundary value problem (5.50) for a sequence of \( m \) values, from \( m = 4 \) to \( m \simeq -0.0904 \). The \( m = 0 \) profile corresponds to the Blasius flow over a flat plate. The case \( m = 1 \) also has specific significance: For \( m = 1 \), the wedge is in fact a vertical wall, like in Figure 5.6c. The inviscid flow is a strain flow whose (dimensional) form is \( u = -\gamma x_1 e_1 + \gamma x_2 e_2 \) for \( x_2 < 0 \). The strain rate \( \gamma \) is a positive constant with units \( 1 \div (\text{time}) \). The only other physical parameter is viscosity \( \nu \), and the unique length determined from \( \gamma \) and \( \nu \) is \( \sqrt{\frac{\nu}{\gamma}} \). This suggests that the boundary layer thickness for \( m = 1 \) is a uniform...
constant proportional to $\sqrt{\frac{u^*}{\gamma}}$. Indeed, for $m = 1$, the similarity variable $\zeta$ in (5.49) reduces to $y$, independent of $s$. The $m = 1$ solution approximates the boundary layer structure near a blunt leading edge, as in Figure 5.10. The $m = 1$ boundary layer flow insinuates itself into models of flow-enhanced condensation onto a solid surface. For instance, “icing” the leading edge of the “wing” in Figure 5.10.

**Precursor of boundary layer separation**

The $m < 0$ velocity profiles $F'(\zeta)$ have an inflexion point $\zeta = \zeta_0 > 0$ so $(F')''(\zeta) > 0$ for $\zeta < \zeta_0$. Significance? The dimensionless vorticity in the boundary layer is approximated by

$$\omega := -\Delta \psi \sim -\epsilon \partial_{yy} \Psi \sim -s^{2m-\frac{1}{2}}F''(\zeta),$$

so the inflexion point $\zeta_0$ marks a minimum of vorticity (strongest concentration of clockwise vortices). As we learn in Problem 5.2, the skin of the body is the source of vorticity in viscous fluid. These vortices are eventually carried downstream from the body, where they aggregate into vortex blobs, doing some complicated space and time dependent “vortex dance”. But even before we reach the trailing edge, here is the usual scenario: The flow just above the boundary layer *decelerates* when you are sufficiently near the trailing edge, and the vortices which stayed near the skin in the accelerating part of the boundary layer tend to “separate” into a vortex layer concentrated away from the skin. In Figure 5.10, the black $x$'s indicate concentration...
of clockwise vortices and the black dots, concentrations of counterclockwise vortices. It is natural to regard the maximum concentration of vortices away from the skin in decelerating boundary layer flows with $-0.0904 < m < 0$ as a precursor of this “separation”.

What is so special about $m = -0.0904$? There is a second branch of (numerical) solutions for $-0.0904 < m < 0$ which merge with the branch in Figure 5.9 at $m = -0.0904$. This second branch of velocity profiles exhibit “backflow” as in Figure 5.11. The situation is analogous to the quadratic equation $(x - x_*)^2 = a$ having two real roots $x = x_* \pm \sqrt{a}$ which merge to $x = x_*$ as $a \to 0^+$, and no real roots for $a < 0$. 

![Figure 5.11](image-url)
Chapter 5. Prantl boundary layer theory
Chapter 6

Modulated oscillations

The intuitive idea of a “modulated oscillation” is a function of time that is nearly periodic over a time interval comparable to a few cycles, but the properties of the oscillation — its local amplitude, local frequency and even its “shape” undergo large cumulative changes after a sufficiently large number of cycles. For instance, recall our first example (1.35), of a harmonic oscillator whose natural frequency is a slowly varying function of time. That is, \( \omega = \frac{1}{1+\epsilon t} \), and limit \( \epsilon \to 0 \) is considered. We examined the exact solution \( x(t, \epsilon) \) of the initial value problem with \( x(0, \epsilon) = 0, \dot{x}(0, \epsilon) = 1 \) as \( \epsilon \to 0 \). Let’s recall what we found: In Exercise 1.4 we computed a two-term expansion in the limit \( \epsilon \to 0 \), \( t > 0 \) fixed. We found that the leading order term \( \dot{x}(t) = \sin t \) is non-uniformly valid as \( t \to \infty \), since the \( O(\epsilon) \) term has a component \( -\frac{\epsilon}{2} t^2 \cos t \) which becomes \( O(1) \) when \( t = O(\sqrt{\epsilon}) \). We see clearly that the naive limit process \( \epsilon \to 0 \) with \( t > 0 \) fixed can’t handle those “large cumulative changes”. In the exact solution we recognized \( \frac{1}{\epsilon} \) as a characteristic time: The local amplitude and frequency of oscillations are functions of “slow time” \( T := \epsilon t \): Take any fixed value \( T = T_0 \) of slow time and look at the “local” limit process \( \epsilon \to 0 \) with \( \tau := t - \frac{T_0}{\epsilon} = O(1) \), corresponding to a “few cycles seen near slow time \( T_0 \)”. In this local limit, the amplitude of oscillations converges to \( \sqrt{1+T_0} \), constant with respect to \( \tau \), and the phase difference between \( \tau = 0 \) and \( \tau \neq 0 \) converges to \( \frac{\tau}{1+T_0} \), indicating the (expected) local frequency \( \frac{1}{1+T_0} \).

The next step in understanding is to recognize that the amplitude \( a \) and
phase \( \zeta \) in the exact solution \( x = a \sin \zeta \) are functions of the slow time,

\[
\begin{align*}
a &= a(T, \epsilon) = \sqrt{\frac{1 + T}{1 - \frac{\epsilon^2}{4}}}, \\
\zeta &= \frac{1}{\epsilon} \theta(T, \epsilon), \quad \theta(T, \epsilon) := \sqrt{1 - \frac{\epsilon^2}{4} \log(1 + T)}.
\end{align*}
\]

This suggests that the solution \( x(t, \epsilon) \) can be constructed asymptotically, starting from the anzatz

\[
x(t, \epsilon) = \text{Im} \left\{ a(T, \epsilon)e^{i \frac{\theta(T, \epsilon)}{\epsilon}} \right\},
\]

reminiscent of the anzatz (2.62) for linear dispersive waves. This is the subject of Problems 6.1 and 6.2.

The anzatz (6.1) is fine for modulations of nearly sinusoidal oscillations, but we also expect to encounter slowly varying non-sinusoidal oscillations. Suppose that the local frequency of oscillations in \( x(t, \epsilon) \) is \( O(1) \) as \( \epsilon \to 0 \), but that its variations, and variations in size and shape of oscillations occur in characteristic time \( \frac{1}{\epsilon} \). A natural generalization of (6.1) is

\[
x(t, \epsilon) = X(\zeta := \frac{\theta(T, \epsilon)}{\epsilon}, T, \epsilon)
\]

where \( X(\zeta, T, \epsilon) \) is \( 2\pi \) periodic in the phase \( \zeta \). This \( 2\pi \) periodicity indicates that the local frequency of oscillation is \( \zeta_t = \theta_T(T, \epsilon) \), an \( O(1) \) function of \( T \). For fixed \( T \), the \( \zeta \)-dependence of \( X(\zeta, T, \epsilon) \) quantifies the local size and shape of oscillations. The explicit \( T := \epsilon t \) dependence indicates changes of oscillation properties in characteristic time \( \frac{1}{\epsilon} \). The anzatz (6.2) is the starting point for the method of two scales — the most straightforward way to construct asymptotic expansions of modulated oscillation solutions to ODE and PDE. But before plunging headlong into its technical content, we survey the

**Physical flavors of modulated oscillations**

In the harmonic oscillator with slowly varying frequency, time translation symmetry is broken, so instead of strict energy conservation, a small amount of work is done on the oscillator per cycle, induced by the slow change of natural frequency. More generally, modulation of oscillations in solutions of
Chapter 6. Modulated oscillations

ODE and PDE can be induced by slow space and or time variations of parameters in the equations. Many other mechanisms induce modulations: In Exercise 1.5, we examined amplitude modulation of oscillatory solutions to the Van-der-Pol ODE induced by a small autonomous, but non-conservative force. Another "usual suspect" is small but periodic variations of parameters in the system. The archetype is resonance — the growing amplitude of harmonic oscillations due to forcing at the natural frequency. Near but not exact resonance usually means beats, with energy going in and out of oscillations over a cycle much longer than the natural period. More generally, if the frequency of the parameter variation is rationally related to an underlying natural frequency in the autonomous dynamics (with the time periodic parameter variation suppressed) we can expect a "generalized resonance" that drives modulations. If we have an autonomous system with many degrees of freedom, these degrees of freedom may "force each other at resonance" and energy is transferred between them in a cyclical pattern in time, like beats. Hence, the nomenclature, "internal resonance" or "internal beats".

Finally, there is homogenization. Historical questions addressed by homogenization are: “What is the effective diffusion tensor of this material with a microscopic, triply periodic variation of a scalar diffusion?” or: “What is the effective dielectric tensor of this crystal, with its periodic molecular lattice?” In these problems, the technical agenda of homogenization is to derive the “course grained” effective diffusion or dielectric tensor from the actual microscopic periodic structure. The physical content of these problems is far from “resonance” and “beats”. But from the perspective of methodology, they are close.

Method of two scales

The vehicle of exposition is a famous example, of weakly non-linear oscillations subject to near resonant forcing. $x(t, \epsilon)$ satisfies the Duffing ODE,

$$
\ddot{x} + x + \epsilon\kappa x^3 = \epsilon a \cos((1 + \epsilon\Omega)t).
$$

(6.3)

Here, $\kappa$, $a$ and $\Omega$ are constants and $\epsilon$ is the gauge parameter. The limit $\epsilon \to 0$ represents weak non-linearity, weak forcing, and small perturbation of the forcing frequency from the natural frequency of linearized oscillations. We can absorb one of the constants $\kappa$, $a$ or $\Omega$, by rescaling $\epsilon$, but we choose not to: The forcing amplitude $\epsilon a$ and frequency deviation $\epsilon\Omega$ are natural “control” parameters, and $\kappa$ positive or negative corresponds to a “hard” or “soft” non-linearity.
Chapter 6. Modulated oscillations

Modulated oscillation solutions to (6.3) as $\epsilon \to 0$ can be represented by a simplified version of (6.2),

$$x(t, \epsilon) = X(\zeta, T := \epsilon t, \epsilon),$$  \hspace{1cm} (6.4)

where

$$\zeta := (1 + \epsilon \Omega)t$$  \hspace{1cm} (6.5)

is the phase of forcing. As in (6.2), $X(\zeta, T, \epsilon)$ is $2\pi$ periodic in $\zeta$. The choice (6.5) of phase $\zeta$ makes sense, because we expect the local period of the weakly non-linear oscillations (in the $\zeta$ variable) is near $2\pi$. We substitute the anzatz (6.4) into the Duffing ODE (6.3). First observe that

$$\dot{x} = (1 + \epsilon \Omega)X_\zeta + \epsilon X_T,$$  \hspace{1cm} (6.6)

so that time differentiation is represented by a directional derivative in $\zeta, T$ space. The ODE (6.3) in terms of $X(\zeta, T, \epsilon)$,

$$\{(1 + \epsilon \Omega)\partial_\zeta + \epsilon \partial_T\}^2 X + X + \epsilon \kappa X^3 = \epsilon a \cos \zeta,$$  \hspace{1cm} (6.7)

is formally a PDE in $\zeta, T$ space. Due to the $2\pi$ periodicity of $X$ in the phase variable $\zeta$, we should think of $\zeta, T$ space as a cylinder, with $(\zeta, T)$ and $(\zeta + 2\pi, T)$ identified with the same point. Figure 6.1 depicts this “modulation cylinder”, with $\zeta$ and $T$ as the circumferential and axial coordinates. On this cylinder, we can “paint” characteristic curves of the directional derivative operator (6.6),

$$\epsilon \zeta - (1 + \epsilon \Omega)T = \text{constant}.$$  \hspace{1cm} (6.8)

We visualize one of these characteristics in Figure 6.1. It “winds up the cylinder slowly”: The change in slow time $T$ in one rotation ($\zeta$ increasing by $2\pi$) is $\frac{2\pi}{1 + \epsilon \Omega}$. The formal character of (6.7) as a PDE on the cylinder is more appearance than reality: The directional derivative (6.6) is, after all, time differentiation along the characteristic curves (6.8).

The $2\pi$ periodicity of $X$ in $\zeta$ informs the real purpose behind this seeming nonsense. Starting from initial values of $X(\zeta, T, \epsilon)$ and $X_\zeta(\zeta, T, \epsilon)$ along $\zeta = 0$ (a vertical line on the cylinder), we can in principle solve the ODE (6.3) along the characteristic curves (6.8) and determine values along $\zeta = 2\pi$. In essence, the ODE (6.3) induces a mapping from $(X(\zeta, T, \epsilon), X_\zeta(\zeta, T, \epsilon))$ at $\zeta = 0$ to $\zeta = 2\pi$. Due to the $2\pi$ periodicity, in $\zeta$, we require $(X(0, T, \epsilon), X_\zeta(0, T, \epsilon))$ to be a fixed point of this “modulation mapping”, and as we shall see, this determines the possible slow time ($T$) dependences of $X(\zeta, T, \epsilon)$. 
Substituting into (6.7) the two-term expansion
\[ X \sim X^0(\zeta, T) + \epsilon X^1(\zeta, T) \] (6.9)
we obtain perturbation equations
\[ X^0_{\zeta\zeta} + X^0 = 0, \] (6.10)
\[ X^1_{\zeta\zeta} + X^1 = -2\omega X^0_{\zeta} - 2X^0_T - \kappa (X^0)^3 + a \cos \zeta. \] (6.11)
The $2\pi$ periodicity of the two-term expansion (6.9) in $\zeta$ implies that $X^0(\zeta, T)$ and $X^1(\zeta, T)$ are $2\pi$ periodic in $\zeta$. The general solution of the leading order equation (6.9) for $X^0(\zeta, T)$ is
\[ X^0 = A(T) \cos \zeta + B(T) \sin \zeta \] (6.12)
where $A(T)$ and $B(T)$ are functions of slow time $T$ independent of $\zeta$. Notice that $X^0(\zeta, T)$ is automatically $2\pi$ periodic in $\zeta$, so the leading order equation provides no information about $A(T)$ and $B(T)$ other than their existence.

Substituting $X^0$ in (6.12) into the right-hand side of (6.11), we have
\[ X^1_{\zeta\zeta} + X^1 = (a + 2\omega A - 2B') \cos \zeta - \kappa (A^3 \cos^3 \zeta + 3AB^2 \cos \zeta \sin^2 \zeta) \\
+ (2\omega B + 2A') \sin \zeta - \kappa (3A^2 B \cos^2 \zeta \sin \zeta + B^3 \sin^3 \zeta). \] (6.13)
In the right-hand side, terms with even and odd symmetry in $\zeta$ are grouped together. The path is now clear: The homogeneous ODE $X^1_{\zeta} + X^1 = 0$ has $2\pi$ periodic solutions $\cos \zeta$ and $\sin \zeta$, so the necessary conditions for $2\pi$ periodic solutions of the inhomogeneous ODE (6.13) are orthogonality of the right-hand side to $\cos \zeta$ and $\sin \zeta$. For instance, orthogonality to $\sin \zeta$ gives

$$ (2\Omega B + 2A') \int_{-\pi}^{\pi} \sin^2 \zeta d\zeta - 3\kappa AB^2 \int_{-\pi}^{\pi} \cos^2 \zeta \sin^2 \zeta d\zeta - \kappa B^3 \int_{-\pi}^{\pi} \sin^4 \zeta d\zeta = 0 $$

or

$$ (2\Omega B + 2A') \left(2\pi \cdot \frac{1}{2}\right) - 3\kappa AB^2 \left(2\pi \cdot \frac{1}{4}\right) - \kappa B^3 \left(2\pi \cdot \frac{3}{8}\right) = 0, $$

and finally,

$$ A' = -\Omega B + \frac{3}{4}\kappa (A^2 + B^2)B. \quad (6.14) $$

Orthogonality to $\cos \zeta$ gives

$$ B' = \frac{a}{2} + \Omega A - \frac{3}{4}\kappa (A^2 + B^2)A. \quad (6.15) $$

In summary, the leading approximation of solutions of the Duffing ODE (6.3) as $\epsilon \to 0$ with $T := \epsilon t = O(1)$ are

$$ x(t, \epsilon) \sim A(\epsilon t) \cos (\zeta := (1 + \epsilon \Omega)t) + B(\epsilon t) \sin \zeta, $$

where $(A(T), B(T))$ is a solution of the modulation ODE (6.14),(6.15). This leading order modulation theory of the Duffing ODE is a first step. Construction of additional terms in the asymptotic expansion of $X(\zeta, T, \epsilon)$ is accompanied by a simultaneous refinement of the modulation ODE (6.14),(6.15). It is more than mechanical repetition of what’s been done so far. Another essential idea enters. But first, its time for a nice physics lesson based on the leading order modulation ODE (6.14),(6.15).
We will see some generic features of weakly non-linear oscillations subject to near resonant forcing. First, the modulation ODE’s are Hamiltonian,

\[ A' = \frac{\partial H}{\partial B}, \quad B' = -\frac{\partial H}{\partial A}, \quad (6.16) \]

where

\[ H = -\frac{a}{2}A - \frac{\Omega}{2}(A^2 + B^2) + \frac{3}{16}\kappa(A^2 + B^2)^2. \quad (6.17) \]

Since \( a, \Omega, \kappa \) are given \textit{constants}, the dynamics are autonomous, and we have an \((A,B)\) phase plane, consisting of level curves of \( H \), parametrized by increasing time. There is a nice way to visualize the phase plane. First, look at \( H \) with \( a = 0 \), corresponding to zero forcing. This \( H \) is a function of the “radial coordinate” \( R := \sqrt{A^2 + B^2} \) only, and the level curves of \( H \) are concentric circles about the origin. If \( \Omega \) and \( \kappa \) have opposite sign, \( H = H(R) \) has one global extremum at \( R = 0 \). The case of \( \Omega, \kappa \) with \textit{same} sign is more interesting: For instance, take \( \Omega \) and \( \kappa \) both positive. The graph of \( H \) as a function of \( a \) and \( b \) looks like the “sombrero” in Figure 6.2a. Now take \( a > 0 \). For \( a \) sufficiently small, we get a “tilted sombrero” as in Figure 6.2b.

It is easy to imagine what the level curves of \( H \) look like. These appear in the phase plane of Figure 6.3. The orientations with increasing \( t \) are figured out by staring at equations (6.16): \( H \) is increasing in \( R := \sqrt{A^2 + B^2} \) for large \( R \), so the closed orbits of large \( R \) have \textit{clockwise} orientation, and the rest of the arrows in Figure 6.3 are done by topological common sense.

The three critical points of \( H \) on the \( A \) axis correspond to the three roots of the cubic equation

\[ \frac{3}{4}\kappa A^3 - \Omega A - a = 0, \quad (6.18) \]

which exist for \( a > 0 \) sufficiently small. The critical points surrounded by small closed orbits in their immediate neighborhood are neutrally stable. The saddle point which is the origin and terminus of homoclinic orbits (darkened in Figure 6.3) is unstable. Physically, the critical points correspond to a periodic solution for \( x(t, \epsilon) \) with amplitude \( |A| \) and whose frequency is the same as that of the forcing, namely \( 1 + \epsilon\Omega \). We can plot the amplitude \( |A| \) as a function of \( \Omega \) for \( \kappa > 0 \) and \( a > 0 \) fixed. First, write (6.18) as

\[ \Omega = \frac{3}{4}\kappa A^2 - \frac{a}{A}, \quad (6.19) \]

which gives \( \Omega \) as a function of \( A \). We start by plotting \( \Omega \) on the horizontal axis, and \( A \) on the vertical. In Figure 6.4a, the dashed curves are the graphs
Chapter 6. Modulated oscillations

\[ \Omega = \frac{3}{2} \kappa A^2 \text{ and } \Omega = -\frac{a}{4}, \]

and the solid curve, obtained by adding \( \Omega \)'s, is the graph of (6.19). The two negative roots \( A \) exist only for 

\[ \Omega > \Omega_c = \left( \frac{3}{2} \right)^{\frac{2}{3}} a^{\frac{1}{3}}. \]  

(6.20)

In our initial discussion, recall that the phase plane in Figure 6.3 based on the “tilted sombrero” was for “a sufficiently small”. From (6.20) we now see that “a sufficiently small” means \( \Omega > \left( \frac{3}{2} \right)^{\frac{2}{3}} a^{\frac{1}{3}} \). For \( \Omega > \Omega_c \), the most negative root of (6.18) corresponds to the unstable saddle point in the phase plane of Figure 6.3. It is marked “u” for “unstable” in Figure 6.4a. The other two roots are neutrally stable, marked “s” in Figure 6.4a. For \( \Omega < \Omega_c \), the single surviving positive root of (6.18) corresponds to a unique, neutrally stable critical point in the \((A, B)\) phase plane. Hence, the \textit{whole} branch of positive \( A \) root in Figure 6.4a is marked “s”. Finally, Figure 6.4b is the graph
of steady amplitude $|A|$ versus $\Omega$, obtained by reflecting the negative $A$ part of Figure 6.4a about the $\Omega$ axis. The portions of the graph corresponding to stable oscillations are solid curves, and the portion corresponding to unstable oscillations is marked by crosses ($\times \times \times$). Figure 6.4b is the traditional “response diagram” for steady, non-linear, forced oscillations. If we redid the whole analysis for negative $\kappa$ instead of $\kappa > 0$ (“soft spring” instead of “hard spring”) the resulting graph of $|A|$ versus $\Omega$ would look like the reflection of Figure 6.4b about the $|A|$ axis.

Here is a piece of physical understanding to go along with the response diagrams. Take the hard spring case with $\kappa > 0$. As $\Omega > 0$ increases, we go away from the resonant frequency 1 of the linearized oscillator. Yet the “s” branch of steady oscillations with $\Omega \sim \frac{3}{4} \kappa A^2$ for $\Omega \gg 1$ persists. What’s that?

Let’s look for a moment at non-linear, but unforced oscillations, with $\kappa > 0$ and $a = 0$. Since there is no forcing, the perturbation $\epsilon \Omega$ of forcing frequency is an irrelevancy. We set $\Omega = 0$. In this case the phase variable $\zeta$ in the modulation representation (6.4) of $x(t, \epsilon)$ is just the time $t$. The modulation Hamiltonian (6.17) reduces to

$$H = \frac{3\kappa}{16} (A^2 + B^2)^2.$$  

The level curves of $H$ are circles, and the slow time required to go once
around the circle of radius $R$ is

$$\frac{2\pi}{\frac{3\kappa}{4} R^2}.$$

If $A(0) = R$, $B(0) = 0$, the solution of the modulation ODE’s (6.14),(6.15) with $a = 0$, $\Omega = 0$ is

$$A(T) = R \cos \left( \frac{3\kappa}{4} R^2 T \right), \quad B(T) = -R \sin \left( \frac{3\kappa}{4} R^2 T \right)$$

and the leading order approximation to $x(t, \epsilon)$ is

$$x^0 = A(T) \cos t + B(t) \sin t = R \cos \left( 1 + \frac{3\epsilon \kappa}{4} R^2 \right) t.$$
Chapter 6. Modulated oscillations

Hence, the natural frequency of the free oscillations of amplitude $R$ is

$$1 + \frac{3\epsilon\kappa}{4} R^2. \quad (6.21)$$

If we want the small forcing to sustain a non-linear oscillation with the amplitude $R$, the forcing frequency $1 + \epsilon\Omega$ should be very near (6.21), hence

$$\Omega \sim \frac{3\kappa}{4} R^2,$$

which is consistent with the $A > 0$ “s” branch of steady response in Figure 6.4b for $\Omega \gg 1$.

Strongly non-linear oscillations and “action”

The modulation theory of strongly non-linear oscillations has distinctive features not fully realized in our previous analysis of linear and weakly non-linear oscillations. The “precept” of fully non-linear modulation theory is “find the right variables!” We’ll see what is meant in the next example, of a garden variety non-linear oscillator subject to small damping: $x(t, \epsilon)$ satisfies ODE

$$\ddot{x} + U'(x) + \epsilon \dot{x} = 0. \quad (6.22)$$

$U(x)$ has a minimum $U(0) = 0$, and no other stationary points for $x$ sufficiently close to zero, as depicted in Figure 6.5. A physically motivated analysis of modulated oscillation solutions to (6.22) has an agenda like this: “The $O(\epsilon)$ damping causes the energy $E := \frac{1}{2} \dot{x}^2 + U(x)$ to decay with characteristic time $\frac{1}{\epsilon}$. Hence we should parametrize the $\epsilon = 0$ orbits in the $x, \dot{x}$ phase plane by $E$, and we’re going to derive a modulation ODE for $E = E(T := \epsilon t)$. And this would work out just fine.
But we’re not going to do that. The methods we’ve developed take us to another place. We substitute into ODE (6.22) the representation of \( x(t, \epsilon) \) as a modulated oscillation,

\[
x = X \left( \zeta := \frac{\theta(T, \epsilon)}{\epsilon}, \ T := \epsilon t, \epsilon \right)
\]

and this results in a “modulation PDE” on the \( \zeta, T \) cylinder,

\[
\omega^2 X_{\zeta\zeta} + \epsilon(\omega T X_\zeta + 2\omega X_{\zeta T}) + \epsilon^2 X_{TT} + U'(X) + \epsilon(\omega X_\zeta + \epsilon X_T) = 0.
\]

Here, \( X(\zeta, T, \epsilon) \) is \( 2\pi \) periodic in phase \( \zeta \), and \( \omega := \theta_T \) is the local frequency of oscillations. Introducing two-term expansions,

\[
X \sim X^0 + \epsilon X^1, \ \theta \sim \theta^0 + \epsilon \theta^1
\]

into (6.24), we obtain perturbation equations

\[
(\omega^0)^2 X^0_{\zeta\zeta} + U'(X^0) = 0 \tag{6.25}
\]

\[
(\omega^0)^2 X^1_{\zeta\zeta} + U''(X^0)X^1 = -2\omega^0 \omega^1 X^0_{\zeta\zeta} - \omega^0 X^1_{\zeta\zeta} - 2\omega^0 X^0_{\zeta T} - \omega^0 X^0_T. \tag{6.26}
\]

Here, \( \omega^0 := \theta^0_T, \ \omega^1 := \theta^1_T \).

It follows from the leading order equation (6.25) that

\[
\frac{1}{2}(\omega^0 X^0_\zeta)^2 + U'(X^0) = E(T) \tag{6.27}
\]

where \( E(T) \) is a function of \( T, \) independent of \( \zeta \). We recognize \( V := \omega^0 X^0_\zeta \) as the leading approximation to velocity \( \dot{x} \), and (6.27) as the good old energy integral of the undamped \( (\epsilon = 0) \) ODE. Figure 6.6 depicts a typical orbit in the traditional \( (X^0, V) \) phase plane corresponding to some positive \( E \). We must choose \( \omega^0 \) to ensure \( 2\pi \) periodicity of \( X^0 \) in \( \zeta \). For the \( V > 0 \) portion of orbit from \( X^0 = a \) to \( X^0 = b \), we have

\[
V = \omega^0 X^0_\zeta = \sqrt{2(E - U(X^0))}.
\]

The increment of phase \( \zeta \) as \( X^0 \) increases by \( dX^0 \) is

\[
d\zeta = \frac{\omega^0}{\sqrt{2(E - U(X^0))}} dX^0.
\]
The elapsed phase difference from $X^0 = a(E)$ to $X^0 = b(E)$ (the two roots of $U(X^0) = E$) is $\pi$, so

$$\pi = \omega^0 \int_{a(E)}^{b(E)} \frac{dX^0}{\sqrt{2(E - U(X^0))}}. \quad (6.28)$$

This determines $\omega^0$ for the orbit. If we parametrize the orbit by $E$, we’d have $\omega^0 = \omega^0(E)$. It seems we are being led to energy after all, and truth be told, energy is an essential part of the story. But now, my young apprentice, there arises a new disturbance in the Force:

The inhomogeneous variational ODE (6.26) for $X^1$ has a $2\pi$ periodic solution only if the right-hand side is orthogonal to the homogeneous solution $X^0_\zeta$. The orthogonality condition can be presented as

$$\frac{d}{dT} \int_0^{2\pi} \omega^0 (X^0_\zeta)^2 d\zeta + \int_0^{2\pi} \omega^0 (X^0_\zeta)^2 d\zeta = 0,$$

or introducing

$$A := \int_0^{2\pi} \omega^0 (X^0_\zeta)^2 d\zeta, \quad (6.29)$$
we have
\[
\frac{dA}{dT} + A = 0. \quad (6.30)
\]
(6.29) is the parametric representation of the line integral
\[
\int_C V dX^0
\]
over the orbit \( C \) in the \( X^0, V = \omega^0 X^0 \) plane in Figure 6.6. This is just the area enclosed by the orbit \( C \). In mechanics, \( A \) is called the action of the orbit \( C \). The action is clearly a monotone increasing function of the energy \( E \), \( A = A(E), A'(E) > 0 \). We see from (6.30) that the action undergoes simple exponential decay in \( T \), so the damping of energy \( E = E(T) \) is given implicitly by
\[
A(E(T)) = A(E(0)) e^{-T}.
\]
We see that modulation theory “chooses” action, rather than energy as the “natural” coordinate for orbits. The selection of action does not happen because of the “\( A \)” term in (6.30) which ultimately arises from the \( \epsilon \dot{x} \) term in the original ODE (6.22). Replace \( \epsilon \dot{x} \) by some general perturbation force \( \epsilon f(x, \dot{x}) \), and this “\( A \)” term is modified. Rather, it is the “\( \frac{dA}{dT} \)” term which always arises, in the same way, independent of the perturbation. Historically, the action as a function of energy arises as a neat way to quantify the temporal frequency of orbits: The area of the orbit \( C \) in Figure 6.6 can be written as
\[
A = A(E) = 2 \int_{a(E)}^{b(E)} \sqrt{2(E - U(X^0))} dX^0. \quad (6.31)
\]
The integrand is \( V \) as a function of \( X^0 \) for the upper half of the orbit in \( V > 0 \). Differentiating (6.31) with respect to \( E \), we have
\[
A'(E) = \sqrt{2} \int_{a(E)}^{b(E)} \frac{dX^0}{\sqrt{E - U(X^0)}}. \quad (6.32)
\]
There are no boundary terms because the integrand in (6.31) vanishes at \( X^0 = a, b \). Comparing (6.28),(6.32), we deduce \( 2\pi = \omega_0 A'(E) \) or
\[
\omega^0 = \frac{2\pi}{A'(E)}. \quad (6.33)
\]
Notice that \( A'(E) \) is the orbit period. Alternatively, we can represent the energy \( E \) of the orbit as a function of area \( A \), \( E = E(A) \). This is consistent
with our premonition that $A$, not $E$ is the “natural” coordinate of orbits. Then, (6.33) is written as

$$\omega^0 = 2\pi E'(A).$$

Having determined $A$ and hence $E$ as functions of the slow time $T$, we can find the leading order term $\theta^0(T)$ of the scaled phase,

$$\theta^0(T) - \theta^0(0) = \int_0^T \omega^0(S)dS = 2\pi \int_0^T \frac{dS}{A'(E(S))}.$$

But we are not quite done: The two-term expansion of the phase is

$$\zeta \sim \frac{\theta^0(T)}{\epsilon} + \theta^1(T)$$

and $\theta^1(T)$ represents an $O(1)$ phase shift. If we really want a leading order approximation to $x(t, \epsilon)$ over characteristic times $O\left(\frac{1}{\epsilon}\right)$, we need to determine $\theta^1(T)$. Rather than pursue this analysis further, we’ll return to the $O(1)$ phase shift later, using an analysis that adopts the action and phase as “proper variables” from the very beginning.

**A primer on non-linear waves**

A famous analysis of Luke and Whitham analyzed modulated wave solutions of the PDE

$$h_{tt} - h_{xx} + U'(h) = 0.$$  \hfill (6.35)

Here, $U(\cdot)$ is the same potential energy function depicted in Figure 6.5. The spatial variable $x$ is accommodated in the anzatz for modulated waves:

$$h(x, t, \epsilon) = H\left(\zeta := \frac{\theta(X, T, \epsilon)}{\epsilon}, \ X := \epsilon x, \ T := \epsilon t, \epsilon\right).$$  \hfill (6.36)

We introduced a slow space variable $X := \epsilon x$ in addition to the slow time $T := \epsilon t$, and the phase $\zeta$ of the waves depends on $X$ as well as $T$. As usual, $H(\zeta, X, T)$ is $2\pi$ periodic in phase $\zeta$, so $k := \theta_X$ is the local wavenumber, and $\omega := -\theta_T$ is the local frequency. (Remember: the minus sign in the definition of $\omega$ is a convention so that linearization of the phase looks like $kx - \omega t$.) Substituting the representation (6.36) into the non-linear wave PDE, we have

$$\omega^2 H_{\zeta\zeta} - \epsilon(\omega_T H_\zeta + 2\omega H_{\zeta T}) + \epsilon^2 H_{tt} - k^2 H_{\zeta\zeta} - \epsilon(k_X H_\zeta + 2k H_{\zeta X}) - \epsilon^2 H_{xx} + U'(H) = 0,$$
or
\[(\omega^2 - k^2)H_{\zeta\zeta} + U'(H) = \epsilon\{\omega_T H_{\zeta} + 2\omega H_{\zeta T} + k_X H_{\zeta} + 2kH_{\zeta X}\} + \epsilon^2(H_{XX} - H_{TT}).\]
The drill is clear: Two-term expansions \(H \sim H^0 + \epsilon H^1, \theta \sim \theta^0 + \epsilon \theta^1\), leading order perturbation equation
\[((\omega^0)^2 - (k^0)^2)H^0_{\zeta\zeta} + U'(H^0) = 0, \quad (6.37)\]
and first-order perturbation equation
\[((\omega^0)^2 - (k^0)^2)H^1_{\zeta\zeta} + U''(H^0)H' = -2(\omega^0 \omega^1 - k^0 k^1)H^0_{\zeta\zeta} + \omega^0 H^0_{\zeta T} + 2\omega^0 H^0_{\zeta X} + k^0 X H^0_{\zeta} + 2k^0 H^0_{\zeta X}. \quad (6.38)\]
The leading order equation implies
\[
\frac{1}{2}((\omega^0)^2 - (k^0)^2)(H^0)^2 + U(H^0) = E(X, T) \quad (6.39)
\]
where \(E(X, T)\) is a function of \(X, T\) independent of \(\epsilon\). For \((\omega^0)^2 - (k^0)^2 > 0\) and \(E > 0\), (6.39) defines closed orbits in the phase plane whose axes are \(H^0\) and \(V := \sqrt{(\omega^0)^2 - (k^0)^2}H^0\). In fact a typical orbit looks exactly like Figure 6.6 with \(H^0\) replacing \(X^0\) on horizontal axis, and the vertical axis is \(V = \sqrt{(\omega^0)^2 - (k^0)^2}H^0\) instead of \(\omega^0 X^0\). Since \(\sqrt{(\omega^0)^2 - (k^0)^2}\) takes over the role of \(\omega^0\), the condition that \(H^0\) is \(2\pi\) periodic in \(\zeta\) is (6.34) with \(\omega^0\) replaced by \(\sqrt{(\omega^0)^2 - (k^0)^2}\), so
\[
\sqrt{(\omega^0)^2 - (k^0)^2} = 2\pi E'(A). \quad (6.40)
\]
Here \(E\) is the constant of \(\zeta\)-integration in (6.39), expressed as a function of the orbit area \(A\). We recognize (6.39) as a dispersion relation for non-linear waves, in which frequency \(\omega^0\) is a function of wavenumber \(k^0\), and also of \(A\), which measures the local amplitude of waves. With \(\omega^0\) an explicit function of \(k^0\) and \(A\), the “conservation of waves”
\[
k^0_T + (\omega^0)_X = 0 \quad (6.41)
\]
(which follows from \(\omega^0 = -\theta^0_T, k^0 = \theta^0_X\)) becomes a PDE in dependent variables \(k^0, A\). The solvability of the inhomogeneous variational ODE (6.38)
subject to $2\pi$ periodicity in $\zeta$ gives a second: First, orthogonality of $H_0^0$ and the right-hand side of (6.38) is expressed as

$$
\left( \omega^0 \int_0^{2\pi} (H_0^0)^2 d\zeta \right)_T + \left( k^0 \int_0^{2\pi} (H_0^0)^2 d\zeta \right) = 0. \quad (6.42)
$$

Here, $\int_0^{2\pi} (H_0^0)^2 d\zeta$ is related to $A$ as in (6.29), with $\sqrt{(\omega^0)^2 - (k^0)^2}$ replacing $\omega^0$, so

$$\int_0^{2\pi} (H_0^0)^2 d\zeta = \frac{A}{\sqrt{(\omega^0)^2 - (k^0)^2}},$$

and (6.42) becomes

$$
\left( \frac{\omega^0 A}{\sqrt{(\omega^0)^2 - (k^0)^2}} \right)_T + \left( \frac{k^0 A}{\sqrt{(\omega^0)^2 - (k^0)^2}} \right)_X = 0,
$$

or with the help of the dispersion relation (6.40),

$$
\left( \frac{\omega^0 A}{E'(A)} \right)_T + \left( \frac{k^0 A}{E'(A)} \right)_X = 0. \quad (6.43)
$$

(6.41) and (6.43) with $\omega^0 = \omega^0(k^0, A)$ given by the dispersion relation (6.40) are a pair of quasilinear PDE in $X, T$ space for $k^0$ an $A$. If hyperbolic, we get two characteristic velocities. If elliptic, the standard initial value problem with $k^0$ and $A$ given functions of $X$ at $T = 0$ is ill posed. In this case, slowly varying wavetrains as assumed in modulation theory are unstable against breaking up into “packets” whose description lies outside of the analysis we’ve done here.

The simplest diagnosis of hyperbolic versus elliptic is by linearization of an equivalent system for $A$ and $\theta^0$ about a uniform wavetrain, with uniform, constant “amplitude” $A$, wavenumber $k^0$, and frequency $\omega^0$. The linearized equations for amplitude and phase perturbations $a(X, T)$ and $\vartheta(X, T)$ are

$$
\omega^0 \vartheta_T + k^0 \vartheta_X = -4\pi^2 E'' E'' a,
$$

$$
\vartheta_{TT} - \vartheta_{XX} + \left( \frac{E''}{E'} - \frac{1}{A} \right) (\omega^0 a_T + k^0 a_X) = 0.
$$

Here, $E' := E'(A)$, $E'' := E''(A)$. These equations are linearizations of the dispersion relation (6.40) and “amplitude” PDE (6.43) respectively. We
can eliminate phase perturbation \( \vartheta \) to obtain a single PDE for amplitude perturbation \( a \):

\[
4\pi^2 E'E''(a_{TT} - a_{XX}) - \left( \frac{E''}{E'} - \frac{1}{A} \right) (\omega^0 \partial_T + k^0 \partial_X)^2 a = 0.
\]

Now look for plane wave solutions for \( a \), proportional to \( e^{i(KX - \omega T)} \). \( \omega \) and \( \kappa \) satisfy dispersion relation

\[
4\pi^2 E'E''(W^2 - K^2) - \left( \frac{E''}{E'} - \frac{1}{A} \right) (k^0 K - \omega^0 W)^2 = 0. \tag{6.44}
\]

Regarding (6.44) as a quadratic equation for \( W \), \( aW^2 + bW + c = 0 \), the discriminant is \( b^2 - 4ac = 64\pi^4 E'^2 E'' \). We have \( A > 0 \), \( E'(A) > 0 \), so the discriminant is positive for \( E''(A) > 0 \). This is the hyperbolic case, in which there are two roots for \( W \) proportional to \( K \), and the two ratios \( \frac{W}{K} \) are characteristic velocities. In this sense we have “two group velocities”.

If \( E''(A) < 0 \), we have the ill-posed elliptic case, and as we said before, uniform waves are expected to break up into packets by a process beyond the scope of this analysis. These two phenomena, “group velocity splitting” and instability are genuinely non-linear phenomena. These effects go away as \( E''(A) \) goes to zero. In the case \( E'(A) = \) uniform constant, the dispersion relation (6.40) becomes amplitude independent, hallmark of the linear case.

**Modulation Lagrangian**

The dispersion relation (6.40) and transport equation (6.43) derive from variation of the **modulation Lagrangian**

\[
L^0(\theta^0_X, \theta^0_T, A) := \sqrt{\theta^0_T^2 - \theta^0_X^2 A} - 2\pi E(A), \tag{6.45}
\]

with respect to \( A \) and \( \theta^0 \). The modulation Lagrangian derives from an asymptotic averaging of the Lagrangian

\[
L := \frac{1}{2}(h_T^2 - h_X^2) - U(h) \tag{6.46}
\]

of the original PDE (6.35). The leading approximation to \( L \) based on the leading approximation \( H^0(\zeta, X, T) \) to \( h \) is

\[
L \sim \frac{1}{2}(\theta^0_T^2 - \theta^0_X^2)(H^0_\zeta)^2 - U(H^0) = \frac{1}{2}(\omega^0 - k^0)^2(H^0_\zeta)^2 - U(H^0). \tag{6.47}
\]
By (6.39),

$$U(H^0) = E(A) - \frac{1}{2}(\omega^0 - k^0)(H^0_\zeta)^2$$

and (6.47) reduces to

$$L \sim (\omega^0 - k^0)(H^0_\zeta)^2 - E(A). \quad (6.48)$$

Now integrate (6.48) with respect to phase $\zeta$ over a period $0 < \zeta < 2\pi$, and use

$$A = \sqrt{\omega^0 - k^0} \int_{-\infty}^{\infty} (H^0_\zeta)^2 d\zeta^*$$

to find

$$\int_0^{2\pi} Ld\zeta \sim \sqrt{\omega^0 - k^0} A - 2\pi E(A)$$

which is $L^0$ in (6.45).

In general, modulation equations that derive from variational ODE or PDE are themselves variational equations. The Lagrangian of the modulation equations is an asymptotic phase average of the Lagrangian associated with the original equations. A “great shortcut to the answer”, when available. We shall see this in

**A primer on homogenization theory**

The simplest example which engages the essential ideas is an elliptic boundary value problem: $u(x, \epsilon)$ satisfies

$$\nabla \cdot \left( D \left( X := \frac{X}{\epsilon} \right) \nabla u \right) = s(x) \quad (6.49)$$

in a region $R$ of $\mathbb{R}^2$, subject given, $\epsilon$-independent boundary values on $\partial R$. Here, $s(x)$ is a given “source” function, and $D(X)$ is a uniformly positive, doubly periodic function whose unit cell is some parallelogram $P$. The characteristic length of oscillations in $D \left( \frac{x}{\epsilon} \right)$ goes to zero as $\epsilon \to 0$. Physically, think of $D \left( \frac{x}{\epsilon} \right)$ as a rapidly oscillating diffusion coefficient. In the limit $\epsilon \to 0$, $u(x, \epsilon)$ converges to a “course grained” field $u^0(x)$, independent of $\epsilon$. $u^0(x)$ assumes the original $\epsilon$-independent boundary values on $\partial R$, and satisfies the elliptic PDE

$$\nabla \cdot (D^0 \nabla u) = s(x). \quad (6.50)$$

*Remember, $A$ is in the area of orbit in $(H^0, V := \sqrt{\omega^0 - k^0} H^0_\zeta)$ phase plane.*
Here, $D^0$ is an effective diffusion tensor, whose cartesian coordinate representation is a $2 \times 2$ symmetric positive definite matrix. The main point of homogenization theory is to derive the effective diffusion tensor $D^0$ from the doubly periodic scalar diffusion $D\left(\frac{x}{\epsilon}\right)$.

First obvious question: Why does $u(x, \epsilon)$ converge positive to an $\epsilon$-independent course grained field $u^0(x)$ as $\epsilon \to 0$? Doesn’t the spatial periodicity of $D\left(\frac{x}{\epsilon}\right)$ induce an oscillatory component in $u(x, \epsilon)$? Well, yes it does, but the amplitude of the oscillatory (“sawtooth”) component is generally $O(\epsilon)$. Look at a one-dimensional example: $u(x, \epsilon)$ satisfies

\[
(D(X)u_x)_x = 0 \text{ in } 0 < x < 1 \\
u(0, \epsilon) = 0, \quad u(1, \epsilon) = U,
\]

where $D(X)$ is a positive function with period 1. We set the source $s(x)$ to zero. The solution is

\[
u(x, \epsilon) = U \frac{\int_0^x \frac{dx'}{D\left(\frac{x'}{\epsilon}\right)}}{\int_0^1 \frac{dx'}{D\left(\frac{x'}{\epsilon}\right)}}.
\]

(6.51)

We represent $\frac{1}{D(X)}$ as a mean value plus fluctuation,

\[
\frac{1}{D(X)} = \left\langle \frac{1}{D} \right\rangle (1 + S'(X)).
\]

(6.52)

Here, brackets denote averaging over $0 < X < 1$,

\[
\left\langle \frac{1}{D} \right\rangle := \int_0^1 dX \frac{X}{D(X)}.
\]

The relative fluctuation $S'(X)$ in (6.52) has zero mean and that’s why we represent it as the derivative of a 1-periodic fluctuation $S(X)$. With the help of (6.52), (6.51) becomes

\[
u = U \frac{x + \epsilon S\left(\frac{x}{\epsilon}\right) - \epsilon S(0)}{1 + \epsilon S\left(\frac{1}{\epsilon}\right) - \epsilon S(0)}
\]

\[
= U x + \epsilon U \left( S\left(\frac{x}{\epsilon}\right) - S\left(\frac{1}{\epsilon}\right) \right) + O(\epsilon^2).
\]

We see that the oscillatory component is asymptotic to $\epsilon US\left(\frac{x}{\epsilon}\right)$ as $\epsilon \to 0$. Hence its contribution to the pointwise value of $u(x, \epsilon)$ vanishes as $\epsilon \to 0$, but
notice that its contribution to $u_x$ is $US'(\frac{\varepsilon}{\epsilon})$, an $O(1)$ oscillation. Similarly, the asymptotic solution of (6.49) has an $O(\epsilon)$ oscillatory component which vanishes as $\epsilon \to 0$, but its gradients remain $O(1)$ as $\epsilon \to 0$. But before we return to the two-dimensional example, what is the effective diffusion coefficient for the one-dimensional case? First, the exact diffusive flux $f := D(X)u_x$ is a uniform constant. With the help of (6.51) we find

$$f = \frac{U}{\int_0^1 \frac{dX'}{D(\frac{X'}{\epsilon})}} = \frac{U}{\langle \frac{1}{D} \rangle} + O(\epsilon).$$

In summary, the flux is proportional to the imposed difference $u(1, \epsilon) - u(0, \epsilon) = 0$, and the coefficient of proportionality $\frac{1}{\langle D \rangle}$ is the effective diffusion coefficient.

The anzatz for the asymptotic solution to (6.49) is

$$u(x, \epsilon) = \langle u \rangle(x, \epsilon) + \epsilon v \left( X := \frac{x}{\epsilon}, x, \epsilon \right).$$

Here, $\langle u \rangle(x, \epsilon)$ and $\epsilon v(X, x, \epsilon)$ are “mean” and “fluctuation” components. The “fluctuation” component $\epsilon v$ is doubly periodic in $X$ with period cell $P$, like $D(X)$, and has zero mean in $X$,

$$\int_D v(X, x, \epsilon) dX = 0. \quad (6.54)$$

The “mean” component converges to the aforementioned coarse grained field $u^0(x)$ in the limit $\epsilon \to 0$. We can introduce the anzatz (6.53) into the PDE (6.49) and formulate perturbation equations based on asymptotic expansions

$$\langle u \rangle \sim u^0 + \epsilon u^1 + \ldots, \quad v \sim v^0 + \epsilon v^1 + \ldots.$$  

The boundary value problem associated with the leading order equation determines $v^0(X, x)$ for $x$ fixed and $X$ in the period cell $P$, as a functional of $\nabla u^0(x)$. A solvability condition of the first-order boundary value problem determines the effective diffusion tensor $D^0$ in (6.50). Details are spelled out in an exercise. Here, we’ll ride the thunderbolt vehicle.†

The PDE (6.49) has Lagrangian density

$$L = \frac{1}{2} D(X) |\nabla u|^2 + s(x)u. \quad (6.55)$$

†In Tibetan Buddhism, the short but perilous ascent to enlightenment.
The one-term expansion of $L$ based on the one-term expansion of $\nabla u$, $\nabla u \sim \nabla u^0 + \nabla_X v^0$, is

$$L^0 = \frac{1}{2} D(X) |\nabla u^0 + \nabla_X v^0|^2 + s(x) u^0.$$  \hfill (6.56)

Here, $\nabla$ and $\nabla_X$ denote gradients with respect to $x$ and $X$. The Euler–Lagrange equation that follows from variation of $v^0$ is

$$\nabla_X \cdot (D(x) \nabla_X v^0) = -\nabla u^0 \cdot \nabla_X D.$$  \hfill (6.57)

(6.57) also follows by substitution of (6.53) into PDE (6.49) and taking limit $\epsilon \to 0$.

We require solution $v^0(X, x)$ which is doubly periodic in $X$ with period cell $P$, and satisfies the zero mean condition (6.54). In practice, the construction of $v^0$ is numerical. Here, we’d like to know, is the problem for $v^0$ well-posed, with a unique solution? Uniqueness modulo an additive constant is easily shown by the energy method: The difference $h$ between two solutions to (6.57) satisfies the homogenous PDE

$$\nabla_X \cdot (D(X) \nabla_X h) = 0.$$  \hfill (6.58)

Multiplying this equation by $h$, integrating in $X$ over period cell $P$, and the usual application of the divergence theorem gives

$$\int_P D(X) |\nabla_X h|^2 dX = \int_{\partial P} D(X) hh \cdot n dS.$$  \hfill (6.58)

By the doubly periodic boundary conditions on $h$, contributions to the line integral from opposite sides of the unit cell cancel, so the right-hand side of
(6.58) is zero. Hence \( \nabla X h \equiv 0 \) in \( P \) and \( h \) is a uniform constant. By the zero mean condition (6.54) the constant is zero.

To determine the effective diffusion tensor \( D^0 \), we need to recognize that (6.57) is really two boundary value problems in disguise: Let \( \omega_1(X) \) and \( \omega_2(X) \) denote the solutions for \( v^0 \) corresponding to \( \nabla u^0 = e_1 \) and \( e_2 \) respectively. Once we’ve constructed numerical solutions of these two unit cell problems, we obtain the solution for \( v^0 \) corresponding to general \( \nabla u^0 \) by superposition

\[
v^0 = (\partial_1 u^0)(x)\omega_1(X) + (\partial_2 u^0)(x)\omega_2(X). \tag{6.59}
\]

Now comes the piece de resistance: Substitute (6.56) for \( v^0 \) into the asymptotic Lagrangian (6.56) and average over \( X \) in unit cell \( P \). We get

\[
\langle L^0 \rangle = \frac{1}{2A} \int_P D(X)[\nabla u^0 + \partial_1 u^0 \nabla X \omega_1 + \partial_2 u^0 \nabla X \omega_2]^2 dX + s(x)u^0
\]

\[
= \frac{1}{2A} \int_P D(X)||((1 + \partial_1 \omega_1)\partial_1 u^0 + \partial_1 \omega_2 \partial_2 u^0)\,e_1 + (\partial_2 \omega_1 \partial_1 u^0 + (1 + \partial_2 \omega_2)\partial_2 u^0)\,e_2|^2 dX + s(x)u^0. \tag{6.60}
\]

Here, \( A \) is the area of \( P \). The integral on the right-hand side can be expressed as a quadratic form in \( \nabla u^0 \), so

\[
\langle L^0 \rangle = \frac{1}{2} \nabla u^0^T D^0 \nabla u^0 + s(x)u^0 \tag{6.61}
\]

where

\[
D^0 := \frac{1}{A} \int_P D(X) \begin{bmatrix} a(X) & c(X) \\ c(X) & b(X) \end{bmatrix} dX \tag{6.62}
\]

with

\[
a := (1 + \partial_1 \omega_1)^2 + (\partial_2 \omega_1)^2 \\
b := (1 + \partial_2 \omega_2)^2 + (\partial_1 \omega_2)^2 \\
c := (1 + \partial_1 \omega_1)\partial_1 \omega_2 + (1 + \partial_2 \omega_2)\partial_2 \omega_1.
\]

\( D^0 \) is a symmetric, positive definite matrix. Since the quadratic form in (6.61) equals the integral in (6.60), we see that it is non-negative for all \( \nabla u^0 \). Can it ever be zero for some \( \nabla u^0 \neq 0 \)? Only if

\[
(1 + \partial_1 \omega_1)\partial_1 u^0 + \partial_1 \omega_2 \partial_2 u^0 = 0, \\
\partial_2 \omega_1 \partial_1 u^0 + (1 + \partial_2 \omega_2)\partial_2 u^0 = 0
\]
Chapter 6. Modulated oscillations

for all \( \mathbf{X} \) in unit cell \( P \) (by inspection of (6.60)) and this in turn implies

\[
(1 + \partial_1 \omega_1)(1 + \partial_1 \omega_2) - \partial_1 \omega_2 \partial_2 \omega_1 = 0 \tag{6.63}
\]

for all \( \mathbf{X} \) in \( P \). Integrating (6.63) in \( \mathbf{X} \) over \( P \), we have

\[
A + \int_P (\partial_1 \omega_1 \partial_2 \omega_2 - \partial_1 \omega_2 \partial_2 \omega_1) d\mathbf{X} = 0.
\]

The integrand is the \( \mathbf{X} \)-divergence of the vector field

\[
\mathbf{F} = \frac{1}{2}(\omega_1 \partial_2 \omega_2 - \omega_2 \partial_2 \omega_1)\mathbf{e}_1 + \frac{1}{2}(\omega_2 \partial_1 \omega_1 - \omega_1 \partial_1 \omega_2)\mathbf{e}_2,
\]

and application of divergence theorem gives

\[
A + \int_{\partial P} \mathbf{F} \cdot \mathbf{n} ds = 0.
\]

The line integral vanishes by periodicity of \( \mathbf{F} \) and we have a contradiction. Hence, the quadratic form \( \frac{1}{2} \nabla \nabla u^0 D^0 \nabla u^0 \) is positive for \( \nabla u^0 \neq 0 \).

The Euler–Lagrange equation which derives from \( u^0 \)-variation of \( \langle L^0 \rangle \) in (6.61) is the effective course-grained diffusion PDE

\[
\nabla \cdot (D^0 \nabla u^0) = s(\mathbf{x}),
\]

and we see that \( D^0 \) in (6.62) is the effective diffusion tensor.