Perturbation Techniques

In this series of lectures, we will like to be introduced to the basics of asymptotic expansions and perturbation techniques. The most elementary application of perturbation techniques is to algebraic equations which depend on a small parameter. We will study three examples, each with increasing complexity. They will clearly demonstrate the ideas, the problems that arise under certain conditions, the reasons for these difficulties (nonuniformities in solutions), and the manner in which these nonuniformities are removed.

Algebraic Equations:

Example 1: Consider the quadratic equation (or, a polynomial of any degree)

\[ x^2 - (3 + 2\varepsilon)x + 2 + \varepsilon = 0, \quad 0 < |\varepsilon| << 1 \]  

(1)

When \(\varepsilon=0\), the equation reduces to the solvable quadratic equation

\[ x^2 - 3x + 2 = (x - 2)(x - 1) = 0. \]  

(2)

Equation (2) is called the reduced equation corresponding to the original problem in equation (1). This equation has the zeros (or, roots) given by

\[ x_1 = 1, \quad x_2 = 2. \]  

(3)

Equation (1) is called the perturbed equation, and it can be thought of as a function \(f(x, \varepsilon)\) whose zeros are being sought. Before we give a general result for the function \(f(x, \varepsilon)\), let us assume that the zeros of the quadratic equation, \(x_i = x_i(\varepsilon)\), \(i = 1, 2\), are analytical functions of the small parameter \(\varepsilon\). Thus, there exist a Taylor series expansion in powers of \(\varepsilon\) which is convergent for every value of \(\varepsilon\) within some radius of convergence \(\varepsilon_0\). Therefore, let

\[ x(\varepsilon) = x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots = \sum_{i=0}^{\infty} \varepsilon^i x_i \]  

(4)

The aim is now to be able to find the coefficients \(x_i, i = 0, 1, 2, 3, 4, \ldots\). Substituting equation (4) into the equation (1) and expanding the resulting expression as an infinite series in \(\varepsilon\) gives

\[ (x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots)^2 - (3 + 2\varepsilon)(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots) + 2 + \varepsilon = 0. \]

Now,
\[(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots)^2 = x_0^2 + 2x_0(\varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots) + (\varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots)^2
\]

\[= x_0^2 + 2\varepsilon x_0 x_1 + \varepsilon^2 (2x_0 x_2 + x_1^2) + \varepsilon^3 (2x_0 x_3 + 2x_1 x_2) + \ldots.\]

and

\[(3 + 2\varepsilon)(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots) = 3x_0 + \varepsilon(3x_1 + 2x_0) + \varepsilon^2 (3x_2 + 2x_1) + \ldots\]

Thus, equation (1) becomes

\[(x_0^2 - 3x_0 + 2) + \varepsilon(2x_0 x_1 - 3x_1 - 2x_0 + 1) + \varepsilon^2 (2x_0 x_2 + x_1^2 - 3x_2 - 2x_1) + \ldots = 0\]  \hspace{1cm} (5)

Setting the coefficient of each power of \(\varepsilon\) to zero separately gives

\[\varepsilon^0 \hspace{1cm} x_0^2 - 3x_0 + 2 = 0\]  \hspace{1cm} (6)

\[\varepsilon^1 \hspace{1cm} 2x_0 x_1 - 3x_1 - 2x_0 + 1 = 0\]  \hspace{1cm} (7)

\[\varepsilon^2 \hspace{1cm} 2x_0 x_2 + x_1^2 - 3x_2 - 2x_1 = 0\]  \hspace{1cm} (8)

Similar equations are obtained at higher orders in \(\varepsilon\). Thus, we get an infinite sequence of algebraic equations which need to be solved sequentially to determine the unknowns \(x_i, i = 0, 1, 2, 3, 4, \ldots\).

Consider the equation (6). It has the two zeros (or, roots)

\[x_{01} = 1, \hspace{1cm} x_{02} = 2.\]  \hspace{1cm} (9)

(Note that equation (6) is the same as the “reduced equation”). Now, consider equation (7). It can be written as

\[x_1 (2x_0 - 3) = 2x_0 - 1\]  \hspace{1cm} (10)

This is a linear nonhomogeneous equation in the unknown variable \(x_1\) with the right hand side being known due to the roots in equation (9). For each of the roots in equation (9), equation (10) can be solved for \(x_1\) provided \((2x_0 - 3) \neq 0\), which is clearly the case. Thus, the two solutions of equation (10) are

\[x_{11} = -1 \hspace{1cm} \text{when} \hspace{1cm} x_{01} = 1, \]

\[x_{12} = 3 \hspace{1cm} \text{when} \hspace{1cm} x_{02} = 2.\]  \hspace{1cm} (11)
Next, consider equation (8) at $O(\epsilon^2)$ written in a slightly different form:

$$x_2(2x_0 - 3) = 2x_1 - x_1^2$$  \hfill (12)

Note again that this is a linear equation for the unknown $x_2$ with nonhomogeneous known term on the right hand side. This equation can also be solved since $(2x_0 - 3) \neq 0$.

The solutions corresponding to the two roots $x_{01}$, $x_{02}$ are

$$x_{21} = 3 \quad \text{when} \quad x_{11} = -1$$

and

$$x_{22} = -3 \quad \text{when} \quad x_{12} = 3.$$  \hfill (13)

Combining the results so far, the perturbation expansions for the two roots of the quadratic polynomial are given by

$$\bar{x}_1 = 1 - \epsilon + 3\epsilon^2 + O(\epsilon^3),$$

$$\bar{x}_2 = 2 + 3\epsilon - 3\epsilon^2 + O(\epsilon^3).$$  \hfill (14)

The correctness of these expansions can be verified by directly writing the two roots of the quadratic equation (1),

$$x = [(3 + 2\epsilon) \pm \sqrt{(3 + 2\epsilon)^2 - 4(2 + \epsilon)}]/2,$$

and then expanding them in a Taylor series for small values of $\epsilon$.

**Remarks:**

1. The power series expansion in equation (4) is also called a **straightforward expansion**.

2. Using this expansion, expanding the nonlinear function $f(x, \epsilon)$ in powers of the small parameter $\epsilon$, and collecting terms of different order in the parameter $\epsilon$ gives an infinite sequence of problems which need to be solved to find the zeros of the function.

3. The first problem in the infinite sequence is identical to the reduced problem obtained by taking the limit of $\epsilon \to 0$. All the subsequent problems are linear equations with the nonhomogeneous terms dependent on the solutions obtained at an earlier level. The coefficient matrices (a scalar quantity in example 1) for the linear problems are always the same functions of the solution of the reduced (lowest order) problem. In fact, the linear coefficient is the partial derivative with respect to the dependent variable $x$ of the original or perturbed equation evaluated at the solution of the reduced problem. Thus, if the linear equation for the first order correction $x_1$ is
solvable (i.e., the coefficient matrix is nonsingular), all equations for higher order corrections are also solvable.

4. Although in the example considered, only three terms in the power series have been explicitly evaluated, one can show that the power series is convergent, i.e., for a fixed \( \varepsilon \) within the radius of convergence, the partial sum \( S_n = \sum_{i=0}^{n} \varepsilon^i x_i \) has a finite limit as \( n \to \infty \). In addition, the limit is the actual zero of the function under consideration. This is really a consequence of the “implicit function theorem” covered in any calculus text.

5. In most applications, one does not proceed to find all the terms in the power series, that is, we usually truncate the series with some number of terms (say, \( N \)). Clearly, we cannot then talk about the convergence of the series as no limit in \( N \to \infty \) can be considered. We are then satisfied if the answer given by the solution with \( N \) terms is close to the true solution, at least for sufficiently small \( \varepsilon \). Thus, one is saying that the difference between the true solution and the \( N \) term solution is at most of \( O(\varepsilon^N) \) for small \( \varepsilon \). This is really the essence of an asymptotic expansion as opposed to a convergent expansion.

We now consider an example which involves a parameter in addition to the small parameter \( \varepsilon \). As we will see, the straightforward expansion will be dependent on this additional parameter, and the validity of the solution will depend on the values of this parameter.

**Example 2:** Consider the quadratic equation

\[
x^2 - (1 + \tau - \varepsilon)x + \tau = 0, \quad 0 < |\varepsilon| << 1, \quad -\infty < \tau < \infty
\]

The corresponding reduced equation (in the limit as \( \varepsilon \to 0 \)) is

\[
x^2 - x(1 + \tau) + \tau = (x - 1)(x - \tau) = 0.
\]

This equation has the roots

\[
\bar{x}_1 = 1, \quad \bar{x}_2 = \tau.
\]

Let us now assume the solutions (roots) of equation (15) as power series

\[
x(\varepsilon) = x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots = \sum_{i=0}^{\infty} \varepsilon^i x_i
\]
Substituting this expansion into equation (15) and expanding the resulting expression in powers of \( \varepsilon \) gives

\[
(x_0 - 1)(x_0 - \tau) + \varepsilon[(2x_0 - 1 - \tau)x_i + x_0] \\
+ \varepsilon^2[(2x_0 - 1 - \tau)x_i^2 + x_i^2 + x_i] + \ldots = 0
\]  

(19)

For this expansion to be identically zero, coefficient of each power of \( \varepsilon \) must identically vanish. Thus, we get an infinite sequence of problems

\[
(x_0 - 1)(x_0 - \tau) = 0
\]

(20)

\[
(2x_0 - 1 - \tau)x_i = -x_0
\]

(21)

\[
(2x_0 - 1 - \tau)x_2 = -(x_i^2 + x_i)
\]

(22)

Equation (20) has the two roots

\[
x_{01} = 1, \quad x_{02} = \tau.
\]

(23)

Now consider equation (21) for the correction \( x_i \) to the roots \( x_{0i}, i = 1, 2 \) of the zeroth order problem. This equation (21) is linear in \( x_i \) and can be solved provided \((2x_0 - 1 - \tau) \neq 0\). Assuming this to be the case, the solutions to equation (21) are given by

\[
x_{11} = -1/(1 - \tau) \quad \text{when} \quad x_{01} = 1
\]

\[
\text{and} \quad x_{12} = \tau/(1 - \tau) \quad \text{when} \quad x_{02} = \tau
\]

(24)

Similarly, equation (22) has the solutions

\[
x_{21} = -\tau/(1 - \tau)^3 \quad \text{when} \quad x_{01} = 1
\]

\[
\text{and} \quad x_{22} = \tau/(1 - \tau)^3 \quad \text{when} \quad x_{02} = \tau
\]

(25)

Combining the terms in the solutions obtained so far, we get the three term expansions for the two roots of the quadratic equation (15) as

\[
\bar{x}_1 = 1 - \varepsilon/(1 - \tau) - \varepsilon^2\tau/(1 - \tau)^3 + O(\varepsilon^3),
\]

\[
\bar{x}_2 = \tau + \varepsilon\tau/(1 - \tau) + \varepsilon^2\tau^2/(1 - \tau)^3 + O(\varepsilon^3)
\]

(26)

Remarks:

1. The series solutions in equations (26) can be shown to be convergent to the true zeros of the quadratic equation in equation (15). If the convergence properties of a series
are independent of the parameter $\tau$, it is said to be uniform. However, here the radius of convergence in $\varepsilon$, $\varepsilon_0$, is a function of the parameter $\tau$. Thus, this series expansion is said to be non-uniform in $\tau$. As is clear from the form of the solutions in equations (26), the solutions breakdown (become “non-uniform”) as $\tau \to 1$ since the higher order terms tend to become unbounded.

2. We first try to find the “region of non-uniformity” of the solution, i.e., the values of the parameter $\tau$ for which the expansion in equations (26) is not useful (i.e., the series with small number of terms does not approximate the true solution reasonably well for at least small values of the parameter $\varepsilon$). Note that the non-uniformity arises when $(\tau-1)$ is small and is $O(\varepsilon^\alpha)$ for some constant $\alpha$. In order to see exactly where this happens, we consider the conditions under which successive terms in the expansion become of the same order (size).

The zeroth and first order terms in the expansion are of the same order when

$$[\varepsilon/(\tau-1)] = O(1) \quad \text{or} \quad (1-\tau) \sim O(\varepsilon) \quad (27)$$

The first order and the second order terms are of the same order when

$$[\varepsilon/(\tau-1)] = O(\varepsilon^2/(1-\tau)^3) \quad \text{or} \quad (1-\tau)^2 \sim O(\varepsilon) \quad (28)$$

Thus, for small $\varepsilon$, the regions of non-uniformity in $\tau$ are dependent on scaling relative to the small parameter $\varepsilon$. There is a region of width of $O(\varepsilon)$ and there is a region of width of $O(\varepsilon^{1/2})$. Both the regions are centered at $\tau=1$ and the region of width of $O(\varepsilon^{1/2})$ is wider than the region of width of $O(\varepsilon)$. Thus, we need to consider the non-uniformity of $O(\varepsilon^{1/2})$.

Now we construct a solution which is valid in the region of width of $O(\varepsilon^{1/2})$ around $\tau=1$. This is called an inner solution and the region around $\tau=1$ is called an inner layer. For this solution, let us introduce a scaling of $\tau$ by

$$(1-\tau) = \varepsilon^{1/2} \sigma \quad (29)$$

where $\sigma$ is independent of $\varepsilon$. Then, equations (15) and (29) give

$$(x-1)(x-1+\varepsilon^{1/2} \sigma) = -\varepsilon \sigma \quad (30)$$

When $\varepsilon=0$, equation (30) reduces to

$$(x-1)^2 = 0 \quad (31)$$
which yields the double root \( x = 1 \). With this in mind, let us assume the solution of equation (30) in a power series in the form

\[
x = 1 + \varepsilon^{1/2} x_1 + \varepsilon x_2 + \varepsilon^{3/2} x_3 + \ldots = 1 + \sum_{n=1}^{\infty} \varepsilon^{n/2} x_n
\]  

(32)

Substituting this solution in equation (30) gives

\[
(\varepsilon^{1/2} x_1 + \ldots)(\varepsilon^{1/2} x_1 + \varepsilon^{1/2} \sigma + \ldots) = -\varepsilon(1 + \varepsilon^{1/2} x_1 + \ldots)
\]  

or \( \varepsilon x_1^2 + \varepsilon \sigma x_1 + \varepsilon + O(\varepsilon^{3/2}) = 0 \)

(33)

At the lowest order in \( \varepsilon \), the reduced equation is

\[
x_1^2 + \sigma x_1 + 1 = 0
\]  

(34)

with roots given by

\[
x_1 = [-\sigma \pm \sqrt{\sigma^2 - 4}] / 2
\]  

(35)

Thus, the solutions of equation (30), upto the second order terms, are given by

\[
x = 1 - \frac{1}{2} \varepsilon^{1/2} (\sigma + \sqrt{\sigma^2 - 4}) + O(\varepsilon)
\]  

\[
x = 1 - \frac{1}{2} \varepsilon^{1/2} (\sigma - \sqrt{\sigma^2 - 4}) + O(\varepsilon)
\]  

(36)

Clearly, these solutions are regular at \( \tau = 1 \) or \( \sigma = 0 \), that is, they do not exhibit any singularity. In terms of the original parameter \( \tau \), these solutions are given by

\[
x_1 = 1 - \frac{1}{2} \varepsilon^{1/2} [\varepsilon^{-1/2} (1 - \tau) + \sqrt{(1-\tau)^2 / \varepsilon - 4}] + O(\varepsilon)
\]  

\[
x_2 = 1 - \frac{1}{2} \varepsilon^{1/2} [\varepsilon^{-1/2} (1 - \tau) - \sqrt{(1-\tau)^2 / \varepsilon - 4}] + O(\varepsilon)
\]  

(37)

This form of the solutions is useful in the neighborhood of \( \tau = 1 \). Far away from \( \tau = 1 \), it is more appropriate to use the straightforward expansions given in equations (26).

We now will give one more example of the types of asymptotic expansions that naturally arise in applications. Before we study this example, let us remind ourselves of the features of the two examples studied. Example 1 was that of a straightforward use of
power series expansions. The reduced problem (limit of small parameter \( \varepsilon \rightarrow 0 \)) possessed two distinct roots and the linear problems for higher order corrections could be sequentially solved without difficulty. In Example 2, the straightforward expansion could again be easily determined except for when the additional parameter was at some critical value. For this value of the second parameter, the higher order terms in the straightforward expansion became singular, or the sequence of linear problems could not be solved for the higher order correction terms. It is interesting to note that at this critical value of the second parameter, the reduced problem has a double root. This is another indicator when the straightforward solutions breakdown and one needs to seek modified procedure for finding solutions. The third situation to be discussed in Example 3 below corresponds to the case when one of the zeros of the function becomes unbounded as the expansion parameter goes to zero. Another way of expressing this is to say that the reduced problem does not give any indication of the existence of some solutions.

**Example 3:** Consider the polynomial

\[ \varepsilon x^2 + x + 1 = 0 \]  

(38)

This equation has to have two roots. Considering the reduced equation however gives

\[ x + 1 = 0 \]  

(39)

which has only one solution. A solution is lost for the reduced problem, which is clearly the case because the reduced problem is a lower degree polynomial compared to the perturbed problem. Such problems are called Singular Perturbation Problems. Now, consider the straightforward expansion for a root in powers of the small parameter \( \varepsilon \). Let

\[ x(\varepsilon) = x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \varepsilon^3 x_3 + \ldots = \sum_{i=0}^{\infty} \varepsilon^i x_i \]  

(40)

Substituting equation (40) in equation (38) gives

\[ \varepsilon(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \ldots)^2 + x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \ldots + 1 = 0 \]

\[ \text{or} \quad x_0 + 1 + \varepsilon(x_1 + x_0^2) + O(\varepsilon^2) = 0 \]  

(41)

Collecting terms of different order in \( \varepsilon \) gives

\[ \varepsilon^0 \quad x_0 + 1 = 0 \]  

(42)

\[ \varepsilon^1 \quad x_1 + x_0^2 = 0 \]

Solving these two equations gives the expansion for one root (upto two terms) as
To understand the essential reason for not being to capture the second root, consider the exact solutions obtained using the quadratic formula:

\[ x = \frac{1}{2\varepsilon} \left[ -1 \mp \sqrt{1-4\varepsilon} \right] \]  

(44)

Let us expand this solution in powers of \(\varepsilon\). To do this, first expand the expression in radical:

\[ \sqrt{1-4\varepsilon} = 1 - 2\varepsilon - 2\varepsilon^2 + \ldots \]  

(45)

Thus, the two roots are:

\[ x = \frac{1}{2\varepsilon} [ -1 + 1 - 2\varepsilon - 2\varepsilon^2 + \ldots ] = -1 - \varepsilon + O(\varepsilon^2) \quad \text{(for +ve sign)} \]

and

\[ x = \frac{1}{2\varepsilon} [ -1 - 1 + 2\varepsilon + 2\varepsilon^2 + \ldots ] = -\frac{1}{\varepsilon} + 1 + \varepsilon + O(\varepsilon^2) \quad \text{(for -ve sign)} \]

(46)

From the second root we can see that the lowest order term is of \(O(\frac{1}{\varepsilon})\) which goes to \(\infty\) as \(\varepsilon \to 0\). The expansion chosen in equation (40) does not contain any term of this type and there cannot capture the root. Thus, the expansion has to be properly chosen to be a useful approximation.

Let us assume that the solution \(x\) is of the form

\[ x(\varepsilon) = y(\varepsilon) / \varepsilon^\nu, \quad \nu > 0 \]  

(47)

The exponent \(\nu\) is unknown at this stage and needs to be determined by using a “dominant term” argument. Substituting the form (47) in equation (40) gives

\[ \varepsilon^{1-2\nu} y^2 + \varepsilon^{-\nu} y + 1 = 0 \]  

(48)

Since \(\nu > 0\), the dominant terms in the expression in equation (48) are the first and the second terms. Thus, they must be captured at the lowest order of approximation of the solution, that is,

\[ \varepsilon^{1-2\nu} y^2 + \varepsilon^{-\nu} y = 0 \]  

(49)
which implies that $1 - 2\nu = -\nu \quad or \quad \nu = 1$. Then, the lowest order approximation equation (49) becomes

$$y^2 + y = 0$$  \hspace{1cm} (50)$$

which has the solutions $y=0$, and $y=-1$. In a more systematic way, with $\nu=1$, equation (48) becomes

$$\epsilon^{-1}y^2 + \epsilon^{-1}y + 1 = 0$$

or

$$y^2 + y + \epsilon = 0$$  \hspace{1cm} (51)$$

We can again use a power series solution for $y(\epsilon)$ of the form

$$y(\epsilon) = y_0 + \epsilon y_1 + \epsilon^2 y_2 + \ldots$$  \hspace{1cm} (52)$$

and proceed to find the constants $y_i, i = 1, 2, 3, \ldots$ in a sequential manner. The equations for zeroth order, first order, and second order approximations are

$$\begin{align*}
\epsilon^0 & \quad y_0^2 + y_0 = 0 \\
\epsilon^1 & \quad (2y_0 + 1)y_1 + 1 = 0 \\
\epsilon^2 & \quad (2y_0 + 1)y_2 + y_1^2 = 0
\end{align*}$$  \hspace{1cm} (53)$$

Solving these equations gives the two roots as

$$\overline{y}_1 = -\epsilon + O(\epsilon^2), \quad \overline{y}_2 = -1 + \epsilon + O(\epsilon^2)$$  \hspace{1cm} (54)$$

The corresponding solutions in terms of the original variable $x$ are

$$\overline{x}_1 = -1 + O(\epsilon), \quad \overline{x}_2 = -1/\epsilon + 1 + O(\epsilon)$$  \hspace{1cm} (55)$$

We can always construct additional terms in the expansions if need arises.