2 Regular perturbation theory

2.1 The Direct Method

Suppose we have a differential equation with initial and/or boundary conditions in which all the variables have been correctly scaled to be dimensionless and there is one other dimensionless coefficient ε that gives the order of magnitude of any term in which it appears. A common situation in applications is one in which ε is small (but not zero) and represents a *perturbation* from a simple problem (hopefully with a known solution) given by setting $\varepsilon = 0$.
A perturbation method is a method of solution that, starting from the known solution of the unperturbed problem ($\varepsilon = 0$), generates in a systematic way an approximate solution to the perturbed problem (which in general we cannot solve exactly).

Example 1.15 (revisited)
It could be that we can't solve (1.31), however we can find an approximate solution by exploiting the fact that we can solve the $\varepsilon = 0$ problem and consider ε to be small.

1) Assume that the solution can be written as a power series in ε, i.e.,

$$u(t) = u(t, \varepsilon) = u_0(t) + \varepsilon u_1(t) + \varepsilon^2 u_2(t) + \cdots .$$

2) Substitute this power series into the problem satisfied by u to obtain

$$(\dot{u}_0 + \varepsilon \dot{u}_1 + \varepsilon^2 \dot{u}_2 + \cdots) = -(u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots) + \varepsilon(u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots)(u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots), \quad (2.1)$$

with initial condition

$$u_0(0) + \varepsilon u_1(0) + \varepsilon^2 u_2(0) + \cdots = 1.$$

Here we use $\dot{u} = du/dt$ for simplicity of notation. The LHS of (2.1) is a power series of the form

$$\sum_{k=0}^{\infty} f_k(t)\varepsilon^k$$

and the RHS is a power series of the form

$$\sum_{k=0}^{\infty} g_k(t)\varepsilon^k$$

and hence

$$\text{LHS} - \text{RHS} = \sum_{k=0}^{\infty} \varepsilon^k (f_k(t) - g_k(t)) = 0,$$

is true for all t and ε.

3) Equate coefficients of ε^k to obtain a series of differential equations with initial conditions

$$(\varepsilon^0) \qquad \dot{u}_0 = -u_0, \qquad u_0(0) = 1, \quad (2.2)$$

$$(\varepsilon^1) \qquad \dot{u}_1 = -u_1 + u_0^2, \qquad u_1(0) = 0, \quad (2.3)$$

$$(\varepsilon^2) \qquad \dot{u}_2 = -u_2 + 2u_0u_1, \qquad u_2(0) = 0, \quad (2.4)$$

etc ...

4) Solve these initial value problems to obtain $u_0,\ u_1,\ u_2, \ldots$. In this case let us solve for terms up to ε^2. From (2.2) we have

$$u_0(t) = e^{-t}$$

and using this in (2.3) gives

$$\begin{aligned}\dot{u_1} + u_1 &= e^{-2t},\ u_1(0) = 0,\\ \Rightarrow\ \frac{d}{dt}(e^t u_1) &= e^t(\dot{u_1} + u_1) = e^{-t},\\ \Rightarrow\ e^t u_1 &= -e^{-t} + A.\end{aligned}$$

Since $u_1(0) = 0$ we have $A = -1$ and hence

$$u_1(t) = e^{-t} - e^{-2t}.$$

Using this and $u_0(t) = e^{-t}$ in (2.4) gives

$$\begin{aligned}\dot{u_2} + u_2 &= 2e^{-2t} - 2e^{-3t},\ u_2(0) = 0,\\ \Rightarrow\ \frac{d}{dt}(e^t u_2) &= e^t(\dot{u_2} + u_2) = 2e^{-t} - 2e^{-2t},\\ \Rightarrow\ e^t u_2 &= -2e^{-t} + e^{-2t} + A.\end{aligned}$$

From our initial data $u_2(0) = 0$ we have $A = 1$ and hence

$$u_2(t) = e^{-t} - 2e^{-2t} + e^{-3t}.$$

5) Choose an approximation by discarding the terms of $O(\varepsilon^k)$ and smaller. In case we take $k = 3$ to obtain

$$\begin{aligned}u_{\text{approx}}(t) &= u_0(t) + \varepsilon u_1(t) + \varepsilon^2 u_2(t)\\ &= e^{-t} + \varepsilon(e^{-t} - e^{-2t}) + \varepsilon^2(e^{-t} - 2e^{-2t} + e^{-3t}).\end{aligned} \tag{2.5}$$

Remark 2.1.

1) *True solution of* (1.31) *can be found by noting that* (1.31) *is a separable first order differential equation,*

$$\left(\frac{1}{-u + \varepsilon u^2}\right)\frac{du}{dt} = 1,$$

Using partial fractions gives

$$\begin{aligned}\left(-\frac{1}{u} + \frac{\varepsilon}{\varepsilon u - 1}\right)\frac{du}{dt} &= 1\\ \Rightarrow\ -\ln u + \ln(\varepsilon u - 1) &= t + C\\ \Rightarrow\ \ln\left(\frac{\varepsilon u - 1}{u}\right) &= t + C\\ \Rightarrow\ \frac{\varepsilon u - 1}{u} &= Ae^t \qquad \textit{where } A = e^C.\end{aligned}$$

Applying $u(0) = 1$ *yields* $A = \varepsilon - 1$

$$\begin{aligned}&\Rightarrow\ u(\varepsilon + (1 - \varepsilon)e^t) = 1\\ &\Rightarrow\ u(t) = \frac{1}{\varepsilon(1 - e^t) + e^t}.\end{aligned}$$

Writing $u(t)$ *in terms of* e^{-t} *so that we can compare it with our approximate solution yields*

$$u(t) = \frac{e^{-t}}{1 + \varepsilon(e^{-t} - 1)}.$$

We note that it is unusual to be able to find a formula for the exact solution of a nonlinear differential equation.

Now let us expand our exact solution in powers of ε, recalling that $(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$ yields

$$\begin{aligned} u(t) &= e^{-t}(1+\varepsilon(e^{-t}-1))^{-1} \\ &= e^{-t}(1-\varepsilon(e^{-t}-1)+\varepsilon^2(e^{-t}-1)^2-\varepsilon^3(e^{-t}-1)^3+\cdots) \\ &= e^{-t}+\varepsilon(e^{-t}-e^{-2t})+\varepsilon^2(e^{-t}-2e^{-2t}+e^{-3t})+\varepsilon^3\cdots \\ &= u_{approx}(t)+\varepsilon^3\cdots . \end{aligned} \tag{2.6}$$

Hence our perturbation method yields the correct approximation up to the ε^3 term.

2) Note that

$$\begin{aligned} u_1(t) &= e^{-t} - e^{-2t}, \\ u_2(t) &= e^{-t} - 2e^{-2t} + e^{-3t} \end{aligned}$$

are order 1 terms for all t and hence do not grow in time. Hence our approximation $u_{approx}(t)$ is valid for all time. The size of the terms are given by the size of ε^k.
Application of this direct method might not yield this in other examples. In particular one frequently sees the appearance of so called secular terms. See next section.

Terminology
The BVP obtained by equating coefficients of ε^0 (i.e. set $\varepsilon = 0$) is the **unperturbed problem**. Its solution is called the *zeroth-order* solution. The BVP obtained by equating coefficients of ε^1 gives the **first order correction** and $u_0 + \varepsilon u$ is the **first order approximation**.
In general, the BVP obtained by equating coefficients of ε^k gives the k**th order correction** u_k and $u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots + \varepsilon^k u_k$ is called the k**th order approximation**.

Recipe for Direct Method

1) Assume that there is a solution in powers of ε.
2) Feed assumed form of solution into BVP (i.e. the differential equation, the initial conditions and the boundary conditions).
3) Equate coefficients of powers of ε which yields a succession of BVPs.
4) Solve as many of these BVPs as is thought necessary for the accuracy required (say up to coefficient of ε^k).
5) Approximate solution is then obtained by discarding the terms not found (i.e. those involving the term ε^{k+1} and higher).

2.2 Poincaré Method for Periodic Solutions

Problems can arise with the direct method if the problem has a periodic solution.
In general, the zeroth order approximation is a solution with a certain period τ_0 (usually 2π as a result of the adopted time scale), but this is NOT the period of the exact solution.
This discrepancy is manifested by the appearance of secular terms in the correction. An example shows this

Example 2.1. *Suppose an object of mass m is displaced a distance y from its equilibrium position so that the force on it, due to a nonlinear spring, is $ky + ay^3$. The equation of motion is*

$$m\frac{d^2y}{d\tau^2} = -ky - ay^3 \qquad \textit{where } a \textit{ small.} \tag{2.7}$$

Let the object be release from rest so that

$$y(0) = A, \quad \frac{dy}{d\tau}(0) = 0. \tag{2.8}$$

Clearly A is an appropriate length scale. Since the unperturbed problem has $a = 0$ and its solution is

$$y(\tau) = A\cos\sqrt{\frac{k}{m}}\tau \qquad \textit{(simple Harmonic motion).}$$

It is oscillatory motion with period $2\pi\sqrt{m/k}$ and so we take our characteristic time scale τ_c to be $\tau_c = \sqrt{m/k}$.

$$\Rightarrow \quad x = \frac{y}{A}, \; t = \frac{\tau}{\sqrt{m/k}}.$$

Rewriting (2.7) and (2.8) in terms of x and t gives

$$\begin{aligned} \Rightarrow \quad & \ddot{x} + x + \varepsilon x^3 = 0, \\ & x(0) = 1, \quad \dot{x}(0) = 0. \end{aligned} \tag{2.9}$$

Here $\varepsilon = aA^2/k$ is a small dimensionless coefficient and (2.9) *is called Duffing's equation. Apply the direct perturbation method to Duffing's equation to find an approximate solution.*

Solution:
Substituting

$$x(t)(= x(t,\varepsilon)) = x_0(t) + \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \cdots$$

into the differential equation and the initial conditions (2.8) we obtain

$$(\ddot{x_0} + \varepsilon\ddot{x}_1 + \varepsilon\ddot{x}_2 + \cdots) + (x_0 + \varepsilon x_1 + \varepsilon x_2 + \cdots) + \varepsilon(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots)^3 = 0,$$

and

$$\begin{aligned} x_0(0) + \varepsilon x_1(0) + \varepsilon x_2(0) + \cdots &= 1, \\ \dot{x_0}(0) + \varepsilon\dot{x}_1(0) + \varepsilon\dot{x}_2(0) + \cdots &= 0. \end{aligned}$$

These equations are true for all $\varepsilon > 0$ and the ODE is true $\forall t > 0$.
We equate coefficients of powers of ε^k on the left and right hand sides of each equation. This procedure yields a succession of differential equation problems for x_0, x_1, x_2, etc. In this case we obtain

$$\ddot{x_0} + x_0 = 0, \quad x_0(0) \; = 1, \quad \dot{x_0}(0) = 0, \tag{2.10}$$

$$\ddot{x}_1 + x_1 + x_0^3 = 0, \quad x_1(0) = 0, \quad \dot{x}_1(0) = 0, \tag{2.11}$$

etc.
We now solve these equations in turn. The solution to (2.10) is $x_0(t) = \cos t$

$$\Rightarrow \; \ddot{x}_1 + x_1 = -\cos^3 t, \quad x_1(0) = 0, \quad \dot{x}_1(0) = 0.$$

(This is an inhomogeneous constant coefficient 2nd order equation.)

$$\begin{aligned} x_1(t) &= x_c + x_p \\ & \left(\begin{array}{l} x_c : \text{ general solution of homogeneous equation} \\ x_p : \text{ any particular solution} \end{array} \right) \\ &= A\cos t + B\sin t + x_p(t). \end{aligned}$$

Noting that

$$\begin{aligned} \cos 3t &= \cos(2t + t) \\ &= \cos 2t\cos t - \sin 2t\sin t \\ &= (\cos^2 t - \sin^2 t)\cos t - 2\sin t\cos t\sin t \\ &= (\cos^2 t - (1 - \cos^2 t))\cos t - 2\cos t(1 - \cos^2 t) \\ &= 4\cos^3 t - 3\cos t \quad \text{Chebyshev polynomial} \end{aligned}$$

we have

$$\begin{aligned} \ddot{x}_1 + x_1 &= -\frac{3}{4}\cos t - \frac{1}{4}\cos 3t, \\ x_p(t) &= C\cos 3t + t(D\cos t + E\sin t). \end{aligned} \tag{2.12}$$

Plugging x_p into the equation for x_1 i.e.

$$\ddot{x}_p + x_p = -\frac{3}{4}\cos t - \frac{1}{4}\cos 3t,$$

we discover that C, D and E need to satisfy certain algebraic equations. In particular

$$C = \frac{1}{32}, \; D = 0, \; E = -\frac{3}{8}.$$

Hence the solution of (2.12) is

$$x_1(t) = A\cos t + B\sin t + \frac{1}{32}\cos 3t - \frac{3}{8}t\sin t.$$

To find A and B we apply initial condition.

$$\Rightarrow \; A = -\frac{1}{32}, \quad B = 0.$$

$$\Rightarrow \; x_1(t) = \frac{1}{32}(\cos 3t - \cos t) - \frac{3t}{8}t\sin t. \tag{2.13}$$

Neglecting terms of ε^2 and higher powers we have

$$\begin{aligned} x_{\text{approx}}(t) &= x_0(t) + \varepsilon x_1(t) \\ &= \cos t + \varepsilon\left[\frac{1}{32}(\cos 3t - \cos t)\right] - \varepsilon\frac{3}{8}t\sin t. \end{aligned}$$

We observe that $|(\cos 3t - \cos t)/32| \leq 1/16$, but the only bound we can put on the remaining term is $|3t\sin t/8| \leq 3t/8$.
The term '$3t\sin t/8$' is an example of a 'secular term'. It grows in time.
Thus $x_0(t)$ is not a good zero order approximation because the first order correction grows in time. This first order correction $\varepsilon x_1(t)$ is not uniformly bounded in terms of ε for all t. We could limit the size of t but this is not useful for looking at oscillatory solutions.
An alternative approach (due to Poincaré) is to scale the independent variable using the (unknown) exact period to get rid of secular term.

Example 2.2 (Poincaré Method). *Same problem as in Example 2.1, initially scaled in the same way to get*

$$\left.\begin{aligned} &\ddot{x} + x + \varepsilon x^3 = 0, \quad t > 0, \\ &x(0) = 1, \; \dot{x}(0) = 0. \end{aligned}\right\} \tag{2.14}$$

Use Poincaré's Method to find an approximate solution to (2.14).

Solution:
Introduce a new term scale by setting

$$s = wt$$

as the independent variable and put

$$w = 1 + \varepsilon w_1 + \varepsilon^2 w_2 + \cdots \tag{2.15}$$

so that the original time scale can be thought of as a zeroth order approximation to the new time scale. The method of Poincaré is to choose the w_k so that each stage of the perturbation procedure any secular term is removed.
We rewrite the equation for $x(s)$ and let a dash denote differentiating with respect to s:

$$\begin{aligned} &w^2 x'' + x + \varepsilon x^3 = 0, \quad s > 0, \\ &x(0) = 1, \quad x'(0) = 0. \end{aligned} \tag{2.16}$$

We now look for

$$x(s) = x_0(s) + \varepsilon x_1(s) + \varepsilon^2 x_2(s) + \cdots,$$
$$\text{and } w = 1 + \varepsilon w_1 + \varepsilon^2 w_2 + \cdots.$$

Plugging these into (2.16) we obtain

$$(1 + \varepsilon w_1 + \varepsilon^2 w_2 + \cdots)^2 (x_0'' + \varepsilon x_1'' + \varepsilon^2 x_2'' + \cdots) + (x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots) + \varepsilon (x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots)^3 = 0,$$

with

$$x(0) = x_0(0) + \varepsilon x_1(0) + \varepsilon^2 x_2(0) + \cdots = 1,$$
$$x'(0) = x_0'(0) + \varepsilon x_1'(0) + \varepsilon^2 x_2'(0) + \cdots = 0.$$

Equating coefficients of ε^k yield a succession of BVPs

$$x_0'' + x_0 = 0, \quad x_0(0) = 1, \quad x_0'(0) = 0, \tag{2.17}$$
$$x_1'' + 2w_1 x_0'' + x_1 + x_0^3 = 0, \quad x_1(0) = 0, \quad x_1'(0) = 0. \tag{2.18}$$

As before (2.17) is solved by

$$x_0(s) = \cos s.$$

To find the first order correction x_1 we need to solve the following :-

$$x_1'' + x_1 = -\cos^3 s + 2w_1 \cos s,$$
$$x_1(0) = 0, \quad x_1'(0) = 0$$

Since

$$\cos 3s = 4\cos^3 s - 3\cos s$$

we have

$$x_1'' + x_1 = \left(2w_1 - \frac{3}{4}\right)\cos s - \frac{1}{4}\cos 3s.$$

The RHS contains a term 'cos' which is a solution of the homogeneous equation. This will lead to a secular term. We use the freedom which arises from the use of $w = 1 + \varepsilon w_1 + \varepsilon^2 w_2 + \cdots$ in order to eliminate secular terms, i.e., choose w_1, w_2, etc in an appropriate way to eliminate the secular terms.
In this case, choosing w_1 such that $2w_1 - 3/4 = 0$ means that the 'bad' term is absent and hence we set $w_1 = 3/8$ to obtain

$$x_1'' + x_1 = -\frac{1}{4}\cos 3s, \qquad x_1(0) = x_1'(0) = 0,$$
$$\Rightarrow \; x_1(s) = \frac{1}{32}(\cos 3s - \cos s).$$

Hence, our first order approximation is

$$x_{\text{approx}}(s) = \cos s + \frac{\varepsilon}{32}(\cos 3s - \cos s),$$

and we note that the first order correction is uniformly small.
We can now return to the original variables.

$$x_{\text{approx}}(\bar{w}t) = \cos \bar{w}t + \frac{\varepsilon}{32}(\cos 3\bar{w}t - \cos \bar{w}t),$$

with

$$\bar{w} = 1 + \frac{3}{8}\varepsilon.$$

Again we see that the first order correction is uniformly small.
We could continue to obtain a second order correction which is uniformly small not containing a secular term by choosing w_2 approximately.

2.3 Validity of Approximations

Let $y_a(t,\varepsilon)$ be an approximation of an exact solution $y(t,\varepsilon)$. The normal requirement for validity is that $y_a(t,\varepsilon)$ should be *uniformly asymptotically valid* for t in some interval I in the following sense.

Definition 2.1 (Uniformly asymptotically valid approximations)**.** *A function $y_a(t,\varepsilon)$ is a uniformly asymptotically valid approximation to a function $y(t,\varepsilon)$ on an interval I as $\varepsilon \to 0$, if the error*

$$E(t,\varepsilon) = y(t,\varepsilon) - y_a(t,\varepsilon)$$

converges to zero as $\varepsilon \to 0$ uniformly for $t \in I$.

What this means is

$$\sup_{t\in I} |E(t,\varepsilon)| \to 0 \text{ as } \varepsilon \to 0,$$

i.e., given $\eta > 0$ (however small) there exists $\varepsilon_0 > 0$ such that $\forall t \in I$,

$$|\varepsilon| < \varepsilon_0 \;\Rightarrow\; |E(t,\varepsilon)| < \eta.$$

While this is alright in theory, but is not very useful in practice, as usually $y(t,\varepsilon)$ is not known, so no expression for $E(t,\varepsilon)$ is available.
One frequently develops theories for bounding $|E(t,\varepsilon)|$ rigorously but this is hard and complicated from an analytical point of view.
However it highlights some difficulties with using approximation methods. Note that the difficulties associated with secular terms are to do with uniformity. We either limit the size of the interval I or use a different approach (e.g. Poincaré).
Something that can be determined is the so called *residual error*. Suppose that (for discussion purposes) our BVP involves a first order differential equation $F(t, y(t), \dot{y}(t), \varepsilon) = 0$ if $y_a(t,\varepsilon)$ is an approximate solution then the residual error is

$$r(t,\varepsilon) = F(r, y_a(t,\varepsilon), \dot{y_a}(t,\varepsilon), \varepsilon).$$

Since we have y_a we can calculate (in principle) the residual.
We say that $y_a(t,\varepsilon)$ satisfies the differential equation **approximately and uniformly** for $t \in I$ provided the residual $r(t,\varepsilon)$ converges uniformly to zero as $\varepsilon \to 0$ for $t \in I$.

Question: If $y_a(t,\varepsilon)$ satisfies the equation approximately and uniformly, then is $y_a(t,\varepsilon)$ a uniformly asymptotically valid approximation?
Answer: Not necessarily. Depends on F, i.e., it depends on the differential equation. This is a difficult question in nonlinear analysis.

Aside: Consider the system of equation $A\underline{x} = \underline{b}$ where the matrix A is given by

$$A = \begin{pmatrix} 1 & 1-\delta \\ 1+\delta & 1 \end{pmatrix}$$

If an approximate solution $\underline{\bar{x}}$ solves $A\underline{\bar{x}} = \underline{b} + \underline{r}$ then setting $\underline{e} = \underline{\bar{x}} - \underline{x}$ gives

$$\begin{array}{l} 0 = A\underline{x} - \underline{b}, \\ \underline{r} = A\underline{\bar{x}} - \underline{b}. \end{array} \quad \Rightarrow \quad A\underline{e} = \underline{r}$$

In this example A is 'nearly' singular since

$$\det A = \delta^2, \;\; \delta \ll 1.$$

It is not the case that $\underline{r}$ small implies $\underline{e}$ is small, since

$$\underline{e} = A^{-1}\underline{r} \;\Rightarrow\; \frac{1}{\delta^2}\begin{bmatrix} 1 & -1+\delta \\ -(1+\delta) & 1 \end{bmatrix}\underline{r}$$

as for $\delta = 10^{-5}$ and

$$\underline{r} = 10^{-2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

we have

$$\underline{e} = 10^{3} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

This is an example of ill conditioning.