Appendix A

Perturbation

A.1 Theory

Here, I provide a general discussion of the perturbation method. It is valid in all contexts where the perturbation theory is applicable, including in quantum mechanics, if the meaning of x are properly given in terms of the quantum mechanical wave-function.[1]

Suppose we have a problem in the form of

$$F(x) = \lambda G(x) \tag{A.1}$$

Here, x is a variable to solve for. Our goal is to find the solution x_s that satisfies

$$F(x_s) = \lambda G(x_s) \tag{A.2}$$

in the vicinity of the "unperturbed solution," or the zero-th order solution, $x^{(0)}$ that satisfies $F(x^{(0)}) = 0$.

If the following conditions are met, then the perturbative approach to solve the above problem is well-defined.

1. λ is any *dimensionless* small number. Small means $|\lambda| \ll 1$. Note that only a dimensionless number can be large or small. In other words, "small" or "large" is a relative concept, not an absolute concept.

[1] In quantum mechanics, the more popular form of the perturbation theory is a more developed one, for which $F(x)$ is assumed to be a certain *specific* function. However, the general discussion here remains valid.

2. $F(x)$ is an analytic function of x, as is its inverse function $(F^{-1}(x))$.[2] Where? In the vicinity of $x^{(0)}$ and x_s. I.e., "where it matters."
   As a practical matter, we require that $F(x)$ is easy to invert. In other words, we require that this step of solution-finding is *easy*:

$$F(x) = const \Rightarrow x = F^{-1}(const)$$

3. $G(x)$ is also an analytic function of x.[3]

Below, we assume that x is a variable, not a function $x(t)$. However, as hinted by footnotes, a generalization to the function case does not pose any fundamental barriers, and it will be left as the reader's work to prove that such a generalization is straightforward.

From the above conditions, it is reasonable to assume that the solution $x_s(\lambda)$ is an analytic function of λ. We will justify this assumption at the end.

The key observation is, then, that, both $F(x_s)$ and $G(x_s)$ are analytic functions of λ, thanks to the chain rule of the differentiation.[4]

$$F(x_s(\lambda)) = \sum_{n=0}^{\infty} f_n \lambda^n \tag{A.3}$$

$$G(x_s(\lambda)) = \sum_{n=0}^{\infty} g_n \lambda^n \tag{A.4}$$

Important terminology: Notice that each term in the summation is of the form a coefficient $\times\lambda^n$. Such a term is said to be $O(\lambda^n)$, or on/of the order of λ^n. Or, it is said to be the n-th order term. As $|\lambda| \ll 1$, by assumption, the n-th order term will be much smaller than the $(n-1)$-th order term.

[2] If x is not a variable, but a function $x(t)$, that is being solved for, then F can be an analytic function of x and its time derivatives (any order), and the theory here will still be applicable, as long as

$$F(x(t), \dot{x}(t), \cdots) = f(t)$$

can be inverted to give

$$x(t) = H(f(t))$$

where H is an analytic function.

[3] If x is not a variable, but a function $x(t)$, that is being solved for, then G can be an analytic function of x and its time derivatives (any order), and the theory here will still be applicable.

[4] Recall that the definition of an analytic function is a function with a well-defined n-th derivative, for any integer $n \geq 1$. Equivalently, a function that has a well-defined Taylor series expansion.

Plugging these two series expansions to Eq. A.2:

$$\sum_{n=0}^{\infty} f_n \lambda^n \;=\; \sum_{n=0}^{\infty} g_n \lambda^{n+1}$$

Now, observe that this equation must be valid for any small λ value, not just one λ value. That is, λ should be considered as a *variable*. Then, it follows that

$$f_0 \;=\; 0, \;\; f_n \;=\; g_{n-1} \;\; (n \geq 1) \tag{A.5}$$

We will see shortly that we can get g_{n-1} from knowing f_{n-1}. Thus, $f_n \;=\; g_{n-1}$ is basically a recursion relation on f_n, the source of the magic that is the perturbation method. The fundamental source of the magic is the fact the coefficient λ is multiplying $G(x)$ in the original equation, of course.

So, let us assume that we know coefficients g_i up to $i = n-1$, for some positive n. This means that we know f_i up to $i = n$. Then, we can set up the following equation.

$$F(x_{s,n}) \;=\; \sum_{i=1}^{n} f_i \lambda^i \tag{A.6}$$

We call this equation **the n-th oder perturbation equation**. This is not an exact equation. Rather, it is an approximation of the exact equation, A.2. More precisely speaking, Eq. A.6 is exact up to $O(\lambda^n)$, but incorrect at higher orders.

Inverting Eq. A.6, we get $x_{s,n} = F^{-1}\left(\sum_{i=1}^{n} f_i \lambda^i\right)$. Despite the obvious math, this is *not yet* the correct solution for Eq. A.6. It is because we are solving a perturbative equation, not an exact equation. Let us instead write $x_{cand,n} = F^{-1}\left(\sum_{i=1}^{n} f_i \lambda^i\right)$, and examine it a little more carefully. Because F^{-1} is an analytic function by our assumption, $x_{cand,n}$ is an analytic function of λ, $x_{cand,n} = \sum_i a_{n,i} \lambda^i$. **The coefficients $a_{n,i}$'s here come from the Taylor expansion of $F^{-1}\left(\sum_{i=1}^{n} f_i \lambda^i\right)$, and they are the meat of the calculation.** But not all $a_{n,i}$'s are trust-worthy. Only those with $i \leq n$ are correct, and $a_{n,i}$'s with $i > n$ should be discarded, when solving the n-th order perturbation equation. Why?

Here is why. Split $x_{cand,n}$ into two terms: $x_{cand,n} = x_{s,n} + x_{s,higher}$ where $x_{s,n} = \sum_{i=0}^{n} a_{n,i} \lambda^i$ and $x_{s,higher} = \sum_{i=n+1}^{\infty} a_{n,i} \lambda^i$. If we evaluate $F(x = x_{cand,n})$, then we see that $x_{s,higher}$ may give rise to terms of $O(\lambda^{n+1})$ and higher, without ever contributing to terms of $O(\lambda^n)$ and lower! But, Eq. A.6 is an approximation to Eq. A.2, up to, and only up to, $O(\lambda^n)$. And so it is *incorrect* to have terms that are higher order than $O(\lambda^n)$ on the left hand side. Physically, that would be "claiming a higher accuracy

than the n-th order perturbation can provide" for the solution. Put another way, at the n-th order perturbation level, all coefficients $a_{n,i}$ with $i > n$ are *incorrect* in general (higher order perturbation solutions, or an exact solution, if available, would show that—observe this fact in the examples provided later) because of the approximate nature of Eq. A.6. However, it is guaranteed that $a_{n,i}$'s are correct for $i \le n$, because Eq. A.6 is correct up to $O(\lambda^n)$. **Therefore, we can drop the index n from $a_{n,i}$ and write it simply as a_i, for $i \le n$.** To conclude, our solution for Eq. A.6 is *definitely not* $x_{s,n} + x_{higher}$, but it is only $x_{s,n}$! **This is the n-th order perturbation solution**:

$$x_{s,n} = \sum_{i=0}^{n} a_i \lambda^i \tag{A.7}$$

The goal of the perturbation calculation at the n-th order is, then, to invert Eq. A.6 to find a_i's ($i = 0 \dots n$). At the n-th order calculation, we only need to find a_n, as all other a_i's ($i = 0 \dots n-1$) are known from the previous order calculation.[5] The following statement, which sums up, and generalizes, the discussion here, should be carefully read and appreciated.

> At the n-th order perturbation approximation, $x_{s,n}$, or any other quantity derived from $x_{s,n}$, is valid only up to $O(\lambda^n)$. The expression for any physical quantity dependent on $x_{s,n}$ should be converted to a Taylor series in λ and terms of higher order than $O(\lambda^n)$ should be discarded. **The n-th order perturbation solution for any physical quantity is always a polynomial of λ, of order n or less (e.g., if a_n happens to vanish).**

In Eq. A.7, if we plug in $n = 0$, then we get $x_{s,0} = a_0$. This is the so-called 0-th order solution that we should obtain by solving $F(x_{s,0}) = 0$. Of course, this is the starting point of the perturbation calculation.

[5] In a more developed perturbation theory that is possible when specific details of F, G are known, e.g. that of quantum mechanics, the perturbation procedure can be written explicitly to do just that—evaluate only a_n (or $x_s^{(n)}$ in page 6) at each step. In a general case, though, one just has to keep in mind that at each order the previous order solution will be repeated and the n-th order correction, which is the only new information, will be added. In practice, this awareness may help make sure that no math mistakes are made. **Most efforts should be focused on correctly obtaining the new term—the n-th order correction—while some basic efforts should be made to double-check lower order terms as well. Observe the solutions given in the example parts of this note, to appreciate this point yourself.**

Here is the magical procedure of the perturbation.

$$
\begin{aligned}
F(x_{s,0}) = 0 \;&\Rightarrow\; x_{s,0} && \\
G(x_{s,0}) \;&\Rightarrow\; g_0 && \text{Eq. A.4} \\
g_0 \;&\Rightarrow\; f_1 && \text{Eq. A.5} \\
f_1 \;&\Rightarrow\; x_{s,1}\textbf{(only } a_1\textbf{is new info)} && \text{Eqs. A.6,A.7} \\
G(x_{s,1}) \;&\Rightarrow\; g_1 && \text{Eq. A.4} \\
g_1 \;&\Rightarrow\; f_2 && \text{Eq. A.5} \\
f_1, f_2 \;&\Rightarrow\; x_{s,2}\textbf{(only } a_2\textbf{is new info)} && \text{Eqs. A.6,A.7} \\
G(x_{s,2}) \;&\Rightarrow\; g_2 && \text{Eq. A.4} \\
&\cdots &&
\end{aligned}
$$

This procedure can be summarized in a simplified form as below.

The following procedure must be fully understood and practiced with ease.

How to solve $F(x) = \lambda G(x)$: $|\lambda| \ll 1$ is a dimensionless number.

$$
\begin{aligned}
F(x) = 0 \;&\Rightarrow\; x_{s,0} && \text{0th order perturbaton solution: } a_0. \\
F(x) = \lambda G(x_{s,0}) \;&\Rightarrow\; x = x_{s,1} && \text{1st order perturbation solution: } a_1 \textbf{ new}. \\
F(x) = \lambda G(x_{s,1}) \;&\Rightarrow\; x = x_{s,2} && \text{2nd order perturbation solution: } a_2 \textbf{ new}. \\
&\cdots && \\
F(x) = \lambda G(x_{s,n-1}) \;&\Rightarrow\; x = x_{s,n} && n\text{-th order perturbation solution: } a_n \textbf{ new}. \\
&\cdots &&
\end{aligned}
$$

$$
\begin{aligned}
x_{s,n} \;&\neq\; F^{-1}(\lambda G(x_{s,n-1})) && \text{Not yet! Need to Taylor-expand it!} \\
x_{s,n} \;&=\; \sum_{i=0}^{n} a_i \lambda^i && \text{Higher order terms are } \textit{illegal}\text{!}
\end{aligned}
$$

$x_{s,n}$ ***must*** **be a polynomial of** λ **... of degree** n**, or less (only if** $a_n = 0$**).**

In theory, this procedure can be carried out to any order, to produce the answer to any accuracy. By taking the limit $n \to \infty$, the exact solution is obtained in a Taylor series form:

$$x_s \;=\; x_{s,\infty} \;=\; \sum_{i=0}^{\infty} a_i \lambda^i$$

This means that the assumption that we made at the start, that x_s is an analytic function of λ, is justified.[6]

In the perturbation theory, it is customary to use the notation

$$x_s^{(n)} = a_n \lambda^n$$

This is the n-th order correction of the perturbing term $(\lambda G(x))$ on the solution x_s. In this notation,

$$x_s = x_s^{(0)} + x_s^{(1)} + x_s^{(2)} + \cdots$$

where all terms are λ dependent except $x_s^{(0)}$. The lowest order non-zero term $x_s^{(n)}$ with $n \geq 1$ is called the **leading order correction** of the perturbation.

A.2 Example

Suppose we have the following perturbation equation to solve with $|\alpha| \ll 1$ and $x \approx 1$.

$$x = 1 + \alpha x^3.$$

With $G(x) = x^3$, $\lambda = \alpha$, and $F(x) = x - 1$, the above equation can be understood as having the perturbation form $F(x) = \lambda G(x)$, and the procedure of the previous page can be directly applied.

$x_{s,0}$	$= 1$	0-th order solution.
$x - 1$	$= \alpha 1^3$	Plug in the 0-th order solution into $G(x)$. This is the 1st order eq.
$x_{s,1}$	$= 1 + \alpha$	1st order soluton. That was easy.
$x - 1$	$= \alpha(1+\alpha)^3$	2nd order equation, obtained by plugging $x_{s,1}$ into $G(x)$.
	$= \alpha + 3\alpha^2 + 3\alpha^3 + \cdots$	Leaving this as is for $x_{s,2}$ would be *incorrect*! (see 5 lines below.)
$x_{s,2}$	$= 1 + \alpha + 3\alpha^2$	2nd order solution. Can keep only up to $O(\alpha^2)$!
$x - 1$	$= \alpha(1 + \alpha + 3\alpha^2)^3$	3rd order equation, obtained by plugging $x_{s,2}$ into $G(x)$.
	$= \alpha + 3\alpha^2 + 12\alpha^3 + \cdots$	Binomial expansion. Higher orders than α^3 are *illegal* for $x_{s,3}$.
$x_{s,3}$	$= 1 + \alpha + 3\alpha^2 + 12\alpha^3$	3rd order solution. Note $12\alpha^3$, not $3\alpha^3$ (see 5 lines above.)
	$\cdots$	We could continue like to any order!

Note that at each step $x_{s,n}$ is $x_{s,n-1}$ plus a correction term of $O(\alpha^n)$. So, one could speed things up if one focuses only on the $O(\alpha^n)$ term in the n-th order equation (or not, if one wishes to double check the previous order solution).

[6] This justification makes sense only if the series converges. If the series does not converge, then the perturbation approach is surely not valid!

A.2.1 Exercise

Obtain the perturbative solutions of the following equations up to the third order in λ. [**Hint for (b):** $\sin(\pi/4+\delta) = \sin(\pi/4)\cos\delta + \cos(\pi/4)\sin\delta$**. Take** δ **to be a small term, and approximate** $\cos\delta$ **and** $\sin\delta$**.**]

(a) $x - 2 = \lambda x^2$. (b) $x = \frac{\pi}{4} + \lambda\sin(x)$. (c) $x - 2 = \lambda\sqrt{1+x}$.

A.3 Another example—air resistance, revisited

Assume that the resistance is $-mkv^2$. Consider a ball falling starting from zero speed.

$$\dot{v} = g - kv^2$$

This equation can be solved perturbatively, assuming a low resistance (small k—but only in a loose sense—read on). In this case, the G function is given as $G(v) = -v^2$.

$$\begin{aligned}
\dot{v} - g &= -kv^2 && \text{EOM to solve}\\
v_{s,0} &= gt && \text{0-th order sol.}\\
\dot{v} - g &= -k(gt)^2 && \text{1-st order eq.}\\
v_{s,1} &= gt - k(gt)^2 t/3 &&\\
&= gt\left(1 - \frac{kgt^2}{3}\right) && \text{1-st order sol.}\\
\dot{v} - g &= -k(gt)^2\left(1 - \frac{kgt^2}{3}\right)^2 && \text{2nd order eq., } \textit{not} \text{ correct yet.}\\
&= -k(gt)^2\left(1 - \frac{2}{3}kgt^2 + \frac{(kgt^2)^2}{9}\right) && \text{The } O(k^3) \text{ term is } \textit{incorrect.}\\
&\approx -k(gt)^2\left(1 - \frac{2}{3}kgt^2\right) && \text{2nd order eq., now correct.}\\
v_{s,2} &= gt\left(1 - \frac{1}{3}kgt^2 + \frac{2}{15}(kgt^2)^2\right) && \text{2nd-order sol. Correct up to } O(k^2)\text{. Good.}\\
\dot{v} - g &= -k(gt)^2\left(1 - \frac{1}{3}kgt^2 + \frac{2}{15}(kgt^2)^2\right)^2 && \text{3rd-order eq. Keep correct terms (up to } O(k^3)\text{) only.}\\
&\approx -k(gt)^2\left(1 - \frac{2}{3}kgt^2 + \frac{17}{45}(kgt^2)^2\right) && \text{Did that. Compare with 4 lines above.}\\
v_{s,3} &= gt\left(1 - \frac{1}{3}kgt^2 + \frac{2}{15}(kgt^2)^2 - \frac{17}{315}(kgt^2)^3\right) && \text{3rd-order sol. Correct up to } O(k^3)\text{. Good!}\\
&\cdots &&
\end{aligned}$$

Here again, notice that, for each $v_{s,n}$, the really new information is only the last term of order $(kgt^2)^n$, while all other terms are repeated from the previous order calculation. That is, $v_{s,n}$ is $v_{s,n-1}$ plus a correction term of the n-th order. This must be the case, always.

What is the true perturbation parameter here? It is *not* k, although, in a loose sense, one may think that it is k. It is actually kgt^2, as the perturbation solution shows. Verify that kgt^2 is dimensionless. This makes much sense. For instance, even if k is a "small" number, the effect of the air resistance definitely cannot be said to be small, if enough time passes so the ball reaches close to the terminal speed situation. Then, the motion is no longer a "free fall plus a small perturbation"! Therefore, it makes sense that, when kgt^2 is large, the above perturbation expansion is not applicable (the mathematics knows this; Higher order terms become large! The series may not converge!). In that case, the large air resistance limit would be applicable, no matter how small the value of k may be in some arbitrary unit.

Lastly, this example is simple enough so that it can be integrated exactly without using the perturbation theory. The answer is

$$v(t) \;=\; v_a \tanh\left(\frac{t}{t_a}\right)$$

where $v_a \equiv \sqrt{g/k}$ and $t_a \equiv 1/\sqrt{kg}$. You are invited to derive this solution, as well as proving that the above perturbation solution and its discussion are in agreement with this exact solution.

In the solution above, the two original parameters of the EOM, g and k, are represented by a new set of two parameters: v_a the velocity scale parameter, and t_a the time scale parameter. The physical meaning of v_a is the terminal speed. The physical meaning of t_a is the cross over time scale below which the motion is basically a free fall motion, albeit perturbed by air resistance, and beyond which the motion becomes *qualitatively different* from a free fall.

Finally, in this example, we assumed that the quadratic air resistance law (air resistance $\propto v^2$) is valid at all speeds. Strictly speaking, this will not be true for low speeds. In the low speed limit[7], the linear air resistance law (air resistance $\propto v$) holds instead. Considering these facts, we might want to restrict the above solution to the cases when the initial velocity is not zero, but a large enough value v_0[8], for which the

[7] Actually, the air resistance ("drag") depends not only on speed, but also on many other factors such as the characteristic dimension, fluid density and viscosity. The "low speed limit" actually means the "low Reynolds number" limit $R \ll 1$. Some discussions of the Reynolds number can found in Gregory's book (Section 4.3), for instance. It should not surprise you that the Reynolds number is a dimensionless number.

[8] Again, the more correct thing to say here is "a v_0 value consistent with a certain range of high Reynolds numbers: $10^3 < R < 10^5$."

quadratic air resistance law is already valid. If we do so, then the above solutions become somewhat complicated. For instance, the exact solution becomes

$$v(t) \;=\; v_a \cdot \frac{(v_a + v_0)e^{2t/t_a} - (v_a - v_0)}{(v_a + v_0)e^{2t/t_a} + (v_a - v_0)}.$$