Chapter 1

Chance and Challenge:
A Brief Review of Homotopy Analysis Method

Shijun Liao
Shanghai Jiao Tong University, Shanghai 200240, China
sjliao@sjtu.edu.cn

A brief review of the homotopy analysis method (HAM) and some of its current advances are described. We emphasize that the introduction of the homotopy, a basic concept in topology, is a milestone of the analytic approximation methods, since it is the homotopy which provides us great freedom and flexibility to choose equation type and solution expression of high-order approximation equations. Besides, the so-called “convergence-control parameter” is a milestone of the HAM, too, since it is the convergence-control parameter that provides us a convenient way to guarantee the convergence of solution series and that differs the HAM from all other analytic approximation methods. Relations of the HAM to the homotopy continuation method and other analytic approximation techniques are briefly described. Some interesting but challenging nonlinear problems are suggested to the HAM community. As pointed out by Georg Cantor (1845–1918), “the essence of mathematics lies entirely in its freedom”. Hopefully, the above-mentioned freedom and great flexibility of the HAM might create some novel ideas and inspire brave, enterprising, young researchers with stimulated imagination to attack them with satisfactory, better results.
1.1. Background

Physical experiment, numerical simulation and analytic (approximation) method are three mainstream tools to investigate nonlinear problems. Without doubt, physical experiment is always the basic approach. However, physical experiments are often expensive and time-consuming. Besides, models for physical experiments are often much smaller than the original ones, but mostly it is very hard to satisfy all similarity criterions. By means of numerical methods, nonlinear equations defined in rather complicated domain can be solved. However, it is difficult to gain numerical solutions of nonlinear problems with singularity and multiple solutions or defined in an infinity domain. By means of analytic (approximation) methods, one can investigate nonlinear problems with singularity and multiple solutions in an infinity interval, but equations should be defined in a simple enough domain. So, physical experiments, numerical simulations and analytic (approximation) methods have their inherent advantages and disadvantages. Therefore, each of them is important and useful for us to better understand nonlinear problems in science and engineering.

In general, exact, closed-form solutions of nonlinear equations are hardly obtained. Perturbation techniques [1–4] are widely used to gain analytic approximations of nonlinear equations. Using perturbation methods, many nonlinear equations are successfully solved, and lots of nonlinear phenomena are understood better. Without doubt, perturbation methods make great contribution to the development of nonlinear science. Perturbation
methods are mostly based on small (or large) physical parameters, called perturbation quantity. Using small/large physical parameters, perturbation methods transfer a nonlinear equation into an infinite number of sub-problems that are mostly linear. Unfortunately, many nonlinear equations do not contain such kind of perturbation quantities at all. More importantly, perturbation approximations often quickly become invalid when the so-called perturbation quantities enlarge. In addition, perturbation techniques are so strongly dependent upon physical small parameters that we have nearly no freedom to choose equation type and solution expression of high-order approximation equations, which are often complicated and thus difficult to solve. Due to these restrictions, perturbation methods are valid mostly for weakly nonlinear problems in general.

On the other side, some non-perturbation methods were proposed long ago. The so-called “Lyapunov’s artificial small-parameter method” [5] can trace back to the famous Russian mathematician Lyapunov (1857–1918), who first rewrote a nonlinear equation

\[ N[u(r,t)] = L_0[u(r,t)] + N_0[u(r,t)] = f(r,t), \]  

where \( r \) and \( t \) denote the spatial and temporal variables, \( u(r,t) \) a unknown function, \( f(r,t) \) a known function, \( L_0 \) and \( N_0 \) are linear and nonlinear operator, respectively, to such a new equation

\[ L_0[u(r,t)] + q N_0[u(r,t)] = f(r,t), \]  

where \( q \) has no physical meaning. Then, Lyapunov regarded \( q \) as a small parameter to gain perturbation approximations

\[ u \approx u_0 + u_1 q + u_2 q^2 + u_3 q^3 + \cdots = u_0 + \sum_{m=1}^{+\infty} u_m q^m, \]  

and finally gained approximation

\[ u \approx u_0 + \sum_{m=1}^{+\infty} u_m, \]  

by setting \( q = 1 \), where

\[ L_0[u_0(r,t)] = f(r,t), \quad L_0[u_1(r,t)] = -N_0[u_0(r,t)], \cdots \]

and so on. It should be emphasized that one has no freedom to choose the linear operator \( L_0 \) in Lyapunov’s artificial small-parameter method: it is exactly the linear part of the whole left-hand side of the original equation \( N[u] = f \), where \( N = L_0 + N_0 \). Thus, when \( L_0 \) is complicated or “singular”
(for example, it does not contain the highest derivative), it is difficult (or even impossible) to solve the high-order approximation equation (1.5). Besides, the convergence of the approximation series (1.4) is not guaranteed in general. Even so, Lyapunov’s excellent work is a milestone of analytic approximation methods, because it is independent of the existence of physical small parameter, even though it first regards $q$ as a “small parameter” but finally enforces it to be 1 that is however not “small” strictly from mathematical viewpoints.

The so-called “Adomian decomposition method” (ADM) [6–8] was developed from the 1970s to the 1990s by George Adomian, the chair of the Center for Applied Mathematics at the University of Georgia, USA. Adomian rewrote (1.1) in the form

$$N[u(r,t)] = L_A[u(r,t)] + N_A[u(r,t)] = f(r,t),$$

(1.6)

where $L_A$ often corresponds to the highest derivative of the equation under consideration, $N_A[u(r,t)]$ gives the left part, respectively. Approximations of the ADM are also given by (1.4), too, where

$$L_A[u_0(r,t)] = f(r,t), \quad L_A[u_m(r,t)] = -A_{m-1}(r,t), \quad m \geq 1,$$

(1.7)

with the so-called Adomial polynomial

$$A_k(r,t) = \frac{1}{k!} \left\{ \frac{\partial^k}{\partial q^k} N_A \left[ \sum_{n=0}^{\infty} u_n(r,t) q^n \right] \right\}_{q=0}.$$  

(1.8)

Since the linear operator $L_A$ is simply the highest derivative of the considered equation, it is convenient to solve the high-order approximation equations (1.7). This is an advantage of the ADM, compared to “Lyapunov’s artificial small-parameter method” [5]. However, the ADM does not provide us freedom to choose the linear operator $L_A$, which is restricted to be related only to the highest derivative. Besides, like “Lyapunov’s artificial small-parameter method” [5], the convergence of the approximation series (1.4) given by the ADM is still not guaranteed.

Essentially, both of the “Lyapunov’s artificial small parameter method” and the “Adomian decomposition method” transfer a nonlinear problem into an infinite number of linear sub-problems, without small physical parameter. However, they have two fundamental restrictions. First, one has no freedom and flexibility to choose the linear operators $L_0$ or $L_A$, since $L_0$ is exactly the linear part of $N$ and $L_A$ corresponds to the highest derivative, respectively. Second, there is no way to guarantee the convergence of the approximation series (1.4). The second ones is more serious, since divergent
approach approximations are mostly useless. Thus, like perturbation methods, the traditional non-perturbation methods (such as Lyapunov’s artificial small parameter method and the ADM) are often valid for weakly nonlinear problems in most cases.

In theory, it is very valuable to develop a new kind of analytic approximation method which should have the following characteristics:

1. It is independent of small physical parameter;
2. It provides us great freedom and flexibility to choose the equation-type and solution expression of high-order approximation equations;
3. It provides us a convenient way to guarantee the convergence of approximation series.

One of such kind of analytic approximation methods, namely the “homotopy analysis method” (HAM) [9–17], was developed by Shijun Liao from 1990s to 2010s, together with contributions of many other researchers in theory and applications. The basic ideas of the HAM with its brief history are described below.

1.2. A brief history of the HAM

The basic ideas of “Lyapunov’s artificial small-parameter method” can be generalized in the frame of the homotopy, a fundamental concept of topology. For a nonlinear equation

\[ N[u(r, t)] = f(r, t), \]  

Liao [9] propose the so-called “homotopy analysis method” (HAM) by using the homotopy, a basic concept in topology:

\[ (1 - q)\mathcal{L} [\varphi(r, t; q) - u_0(r, t)] = c_0 \quad q \quad H(r, t) \quad (N[\varphi(r, t; q)] - f(r, t)), \]  

where \( \mathcal{L} \) is an auxiliary linear operator with the property \( \mathcal{L}[0] = 0 \), \( N \) is the nonlinear operator related to the original equation (1.9), \( q \in [0, 1] \) is the embedding parameter in topology (called the homotopy parameter), \( \varphi(r, t; q) \) is the solution of (1.10) for \( q \in [0, 1] \), \( u_0(r, t) \) is an initial guess, \( c_0 \neq 0 \) is the so-called “convergence-control parameter”, and \( H(r, t) \) is an auxiliary function that is non-zero almost everywhere, respectively. Note that, in the frame of the homotopy, we have great freedom to choose the auxiliary linear operator \( \mathcal{L} \), the initial guess \( u_0(r, t) \), the auxiliary function \( H(r, t) \), and the value of the convergence-control parameter \( c_0 \).
When $q = 0$, due to the property $L[0] = 0$, we have from (1.10) the solution
\[ \varphi(r, t; 0) = u_0(r, t). \] (1.11)

When $q = 1$, since $c_0 \neq 0$ and $H(r, t) \neq 0$ almost everywhere, Eq. (1.10) is equivalent to the original nonlinear equation (1.9) so that we have
\[ \varphi(r, t; 1) = u(r, t), \] (1.12)
where $u(r, t)$ is the solution of the original equation (1.9). Thus, as the homotopy parameter $q$ increases from 0 to 1, the solution $\varphi(r, t; q)$ of Eq. (1.10) varies (or deforms) continuously from the initial guess $u_0(r, t)$ to the solution $u(r, t)$ of the original equation (1.9). For this sake, Eq. (1.10) is called the zeroth-order deformation equation.

Here, it must be emphasized once again that we have great freedom and flexibility to choose the auxiliary linear operator $L$, the auxiliary function $H(r, t)$, and especially the value of the convergence control parameter $c_0$ in the zeroth-order deformation equation (1.10). In other words, the solution $\varphi(r, t; q)$ of the zeroth-order deformation equation (1.10) is also dependent upon all\(^4\) of the auxiliary linear operator $L$, the auxiliary function $H(r, t)$ and the convergence-control parameter $c_0$ as a whole, even though they have no physical meanings. This is a key point of the HAM, which we will discuss in details later. Assume that $L, H(r, t)$ and $c_0$ are properly chosen so that the solution $\varphi(r, t; q)$ of the zeroth-order deformation equation (1.10) always exists for $q \in (0, 1)$ and besides it is analytic at $q = 0$, and that the Maclaurin series of $\varphi(r, t; q)$ with respect to $q$, i.e.
\[ \varphi(r, t; q) = u_0(r, t) + \sum_{m=1}^{+\infty} u_m(r, t) q^m \] (1.13)
converges at $q = 1$. Then, due to (1.12), we have the approximation series
\[ u(r, t) = u_0(r, t) + \sum_{m=1}^{+\infty} u_m(r, t). \] (1.14)

Substituting the series (1.13) into the zeroth-order deformation equation (1.10) and equating the like-power of $q$, we have the high-order approximation equations for $u_m(r, t)$, called the $m$th-order deformation equation
\[ L[u_m(r, t) - \chi_m u_{m-1}(r, t)] = c_0 H(r, t) R_{m-1}(r, t), \] (1.15)
\(^4\)More strictly, $\varphi(r, t; q)$ should be replaced by $\varphi(r, t; q, L, H(r, t), c_0)$. Only for the sake of simplicity, we use here $\varphi(r, t; q)$, but should always keep this point in mind.
where
\[
R_k(r, t) = \frac{1}{k!} \left\{ \frac{\partial^k}{\partial q^k} \left( N \left[ \sum_{n=0}^{+\infty} (r, t) q^n \right] - f(r, t) \right) \right\} \bigg|_{q=0},
\]
with the definition
\[
\chi_k = \begin{cases} 
0, & \text{when } k \leq 1, \\
1, & \text{when } k \geq 2.
\end{cases}
\]

For various types of nonlinear equations, it is easy and straightforward to use the theorems proved in Chapter 4 of Liao’s book \([11]\) to calculate the term \(R_k(r, t)\) of the high-order deformation equation (1.15).

It should be emphasized that the HAM provides us great freedom and flexibility to choose the auxiliary linear operator \(\mathcal{L}\) and the initial guess \(u_0\). Thus, different from all other analytic methods, the HAM provides us great freedom and flexibility to choose the equation type and solution expression of the high-order deformation equation (1.15) so that its solution can often be gained without great difficulty. Notice that “the essence of mathematics lies entirely in its freedom”, as pointed out by Georg Cantor (1845–1918). More importantly, the high-order deformation equation (1.15) contains the convergence-control parameter \(c_0\), and the HAM provides great freedom to choose the value of \(c_0\). Mathematically, it has been proved that the convergence-control parameter \(c_0\) can adjust and control the convergence region and ratio of the approximation series (1.14). For details, please refer to Liao [10, 12, 13] and especially § 5.2 to § 5.4 of his book [11]. So, unlike all other analytic approximation methods, the convergence-control parameter \(c_0\) of the HAM provides us a convenient way to guarantee the convergence of the approximation series (1.14). In fact, it is the convergence-control parameter \(c_0\) that differs the HAM from all other analytic methods.

At the \(m\)th-order of approximation, the optimal value of the convergence-control parameter \(c_0\) can be determined by the minimum of residual square of the original governing equation, i.e.
\[
\frac{d\mathcal{E}_m}{dc_0} = 0,
\]
where
\[
\mathcal{E}_m = \int_{\Omega} \left\{ N \left[ \sum_{n=0}^{m} u_n(r, t) \right] - f(r, t) \right\}^2 d\Omega.
\]

Besides, it has been proved by Liao [16] that a homotopy series solution (1.14) must be one of solutions of considered equation, as long as it is
convergent. In other words, for an arbitrary convergence-control parameter $c_0 \in \mathbb{R}_c$, where

$$
\mathbb{R}_c = \left\{ c_0 : \lim_{m \to +\infty} \mathcal{E}_m(c_0) \to 0 \right\}
$$

is an interval, the solution series (1.14) is convergent to the true solution of the original equation (1.9). For details, please refer to Liao [16] and Chapter 3 of his book [11].

In summary, the HAM has the following advantages:

(a) it is independent of any small/large physical parameters;
(b) it provides us great freedom and large flexibility to choose equation type and solution expression of linear high-order approximation equations;
(c) it provides us a convenient way to guarantee the convergence of approximation series.

In this way, nearly all restrictions and limitations of the traditional non-perturbation methods (such as Lyapunov’s artificial small parameter method [5], the Adomian decomposition method [6–8], the $\delta$-expansion method [18] and so on) can be overcome by means of the HAM.

Besides, it has been generally proved [10, 12, 13] that the Lyapunov’s artificial small parameter method [5], the Adomian decomposition method [6–8] and the $\delta$-expansion method [18] are only special cases of the HAM for some specially chosen auxiliary linear operator $\mathcal{L}$ and convergence-control parameter $c_0$. Especially, the so-called “homotopy perturbation method” (HPM) [19] proposed by Jihuan He in 1998 (six years later after Liao [9] proposed the early HAM in 1992) was only a special case of the HAM when $c_0 = -1$, and thus has “nothing new except its name” [20]. Some results given by the HPM are divergent even in the whole interval except the given initial/boundary conditions, and thus “it is very important to investigate the convergence of approximation series, otherwise one might get useless results”, as pointed out by Liang and Jeffrey [21]. For details, see § 6.2 of Liao’s book [11]. Thus, the HAM is more general in theory and widely valid in practice for more of nonlinear problems than other analytic approximation techniques.

In calculus, the famous Euler transform is often used to accelerate convergence of a series or to make a divergent series convergent. It is interesting that one can derive the Euler transform in the frame of the HAM, and give a similar but more general transform (called the generalized Euler transform), as shown in Chapter 5 of Liao’s book [11]. This provides us a
theoretical cornerstone for the validity and generality of the HAM.

The introduction of the so-called “convergence-control parameter” $c_0$ in the zeroth-order deformation equation (1.10) is a milestone for the HAM. From physical viewpoint, the “convergence-control parameter” $c_0$ has no physical meanings so that convergent series of solution given by the HAM must be independent of $c_0$. This is indeed true: there exists such a region $R_{c_0}$ that, for arbitrary $c_0 \in R_{c_0}$, the HAM series converges to the true solution of the original equation (1.9), as illustrated by Liao [10, 11]. However, if $c_0 \notin R_{c_0}$, the solution series diverges! So, from a mathematical viewpoint, the “convergence-control parameter” is a key point of the HAM, which provides us a convenient way to guarantee the convergence of the solution series. In fact, it is the so-called “convergence-control parameter” that differs the HAM from all other analytic approximation methods.

The introduction of the basic concept homotopy in topology is also a milestone of the analytic approximation methods for nonlinear problems. It is the homotopy that provides us great freedom and large flexibility to choose the auxiliary linear operator $L$ and initial guess $u_0$ in the zeroth-order deformation equation (1.10), which determine the equation type and solution expression of the high-order deformation equations (1.15). Besides, it is the homotopy that provides us the freedom to introduce the so-called “convergence-control parameter” $c_0$ in (1.10), which becomes now a cornerstone of the HAM. Note that it is impossible to introduce such kind of “convergence-control parameter” in the frame of perturbation techniques and the traditional non-perturbation methods (such as Lyapunov’s artificial small parameter, Adomian decomposition method and so on).

The freedom on the choice of the auxiliary linear operator $L$ is so large that the second-order nonlinear Gelfand equation can be solved conveniently (with good agreement with numerical results) in the frame of the HAM even by means of a forth-order auxiliary linear operator (for two dimensional Gelfand equation) or a sixth-order auxiliary linear operator (for three dimensional Gelfand equation), respectively, as illustrated by Liao [14]. Although it is true that the auxiliary linear operator (with the same highest order of derivative as that of considered problem) can be chosen straightforwardly in most cases, such kind of freedom of the HAM should be taken into account sufficiently by the HAM community when necessary, especially for some valuable but challenging problems (some of them are suggested below in § 1.5).

In addition, by means of the above-mentioned freedom of the HAM, the convergence of approximation solution can be greatly accelerated in the
frame of the HAM by means of the iteration, the so-called homotopy-Padé
technique and so on. For details, please refer to § 2.3.5 to § 2.3.7 of Liao’s

Indeed, “the essence of mathematics lies entirely in its freedom”, as
pointed out by Georg Cantor (1845–1918).

Such kind of great freedom of the HAM should provide us great possi-
bility to solve some open questions. One of them is described below. The
solution of the high-order deformation equation (1.15) can be expressed in
the form

\[ u_m(r, t) = -\chi_m u_{m-1}(r, t) + L^{-1} \left[ c_0 H(r, t) R_{m-1}(r, t) \right], \]  

(1.21)

where \( L^{-1} \) is the inverse operator of \( L \). For a few auxiliary linear operator
\( L \), its inverse operator is simple. However, in most cases, it is not straight-
forward to solve the above linear differential equation. Can we directly
choose (or define) the inverse auxiliary linear operator \( L^{-1} \) so as to solve
(1.15) conveniently? This is possible in the frame of the HAM, since in
theory the HAM provides us great freedom and large flexibility to choose
the auxiliary linear operator \( L \). If successful, it would be rather efficient
and convenient to solve the high-order deformation equation (1.15). This is
an interesting but open question for the HAM community, which deserves
to be studied in details. Note that some interesting problems are suggested
in § 1.5.

1.3. Some advances of the HAM

Since 1992 when Liao [9] proposed the early HAM, the HAM has been
developing greatly in theory and applications, due to the contributions of
many researchers in dozens of countries. Unfortunately, it is impossible to
describe all of these advances in details in this brief review, and even in this
book. In fact, the HAM has been successfully applied to numerous, various
types of nonlinear problems in science, engineering and finance. So, we had
to focus on a rather small port of these advances here.

1.3.1. Generalized zeroth-order deformation equation

The starting point of the use of the HAM is to construct the so-called zeroth-
order deformation equation, which builds a connection (i.e. a continuous
mapping/deformation) between a given nonlinear problem and a relatively
much simpler linear ones. So, the zeroth-order deformation equation is a
base of the HAM.
Given a nonlinear equation, we have great freedom and large flexibility in the frame of the HAM to construct the so-called zeroth-order deformation equation using the concept homotopy in topology. Especially, the convergence-control parameter $c_0$ plays an important role in the frame of the HAM. So, it is natural to enhance the ability of the so-called “convergence control” by means of introducing more such kind of auxiliary parameters. Due to the above-mentioned freedom and flexibility of the HAM, there are numerous approaches to do so. For example, we can construct such a kind of zeroth-order deformation equation with $K + 1$ convergence-control parameters:

$$
(1 - q)L[\varphi(r,t;q) - u_0(r,t)] = \left(\sum_{n=0}^{K} c_n q^{n+1}\right) H(r,t) \{N[\varphi(r,t;q)] - f(r,t)\},
$$

(1.22)

where $\varphi(r,t;q)$ is the solution, $N$ is a nonlinear operator related to an original problem $N[u(r,t)] = f(r,t)$, $q \in [0, 1]$ is the homotopy parameter, $u_0$ is an initial guess, $L$ is an auxiliary linear operator, $H(r,t)$ is an auxiliary function which is nonzero almost everywhere, and

$$
c = \{c_0, c_1, \ldots, c_K\}
$$

is a vector of $(K + 1)$ non-zero convergence-control parameters, respectively. Note that, when $K = 0$, it gives exactly the zeroth-order deformation equation (1.10).

The corresponding high-order deformation equation reads

$$
L[u_m(r,t) - \chi_m u_{m-1}(r,t)] = H(r,t) \sum_{n=0}^{\min\{m-1, K\}} c_n R_{m-1-n}(r,t),
$$

(1.23)

where $R_n(r,t)$ and $\chi_n$ are defined by the same formulas (1.16) and (1.17), respectively. When $K = 0$, the above high-order deformation equation (1.23) is exactly the same as (1.15). At the $m$th-order of approximation, the optimal convergence-control parameters are determined by the minimum of the residual square of the original equation, i.e.

$$
\frac{d\mathcal{E}_m}{dc_n} = 0, \quad 0 \leq n \leq \min\{m - 1, K\},
$$

(1.24)

where $\mathcal{E}_m$ is defined by (1.19). For details, please refer to Chapter 4 of Liao’s book [11]. When $K \to +\infty$, it is exactly the so-called “optimal homotopy asymptotic method” [22]. So, the “optimal homotopy asymptotic method”...
In theory, the more the convergence-control parameters, the larger the ability to control the convergence of the HAM series. However, it is found [16] that much more CPU times is needed in practice when more convergence-control parameters are used. In most cases, one optimal convergence-control parameter is good enough to gain convergent results by means of the HAM. Considering the computational efficiency, one up to three convergence-control parameters are generally suggested in the frame of the HAM. For details, please refer to § 2.3.3, § 2.3.4, § 4.6.1 and Chapter 3 of Liao’s book [16].

It should be emphasized once again that, in the frame of the homotopy in topology, we have rather great freedom and large flexibility to construct the so-called zeroth-order deformation equation. In theory, given a nonlinear equation $N[u(r, t)] = f(r, t)$, we can always properly choose an initial guess $u_0(r, t)$ and an auxiliary linear operator $L$ to construct such a zeroth-order deformation equation in a rather general form

$$A[u_0(r, t), L, \varphi(r, t; q), c; q] = 0 \quad (1.25)$$

that it holds

$$\varphi(r, t; 0) = u_0(r, t), \quad \text{when } q = 0, \quad (1.26)$$

and

$$\varphi(r, t; 1) = u(r, t), \quad \text{when } q = 1, \quad (1.27)$$

i.e., when $q = 1$ the zeroth-order deformation equation (1.25) is equivalent to the original nonlinear equation $N[u(r, t)] = f(r, t)$. Using the theorems given in Chapter 4 of Liao’s book [11], it is easy to gain the corresponding high-order deformation equations. Here,

$$c = \{c_0, c_1, \cdots, c_K\}$$

is a vector of convergence-control parameters, whose optimal values are determined by the minimum of residual square of the original equation. Note that (1.25) is rather general: the zeroth-order deformation equations (1.10) and (1.22), and even Eq. (1.2) for Lyapunov’s artificial small parameter method, are only special cases of (1.25). Some commonly used zeroth-order deformation equations are described in § 4.3 of Liao’s book [11] as special cases of the generalized zeroth-order deformation equation (1.25).

In theory, there are an infinite number of different ways to construct a zeroth-order deformation equation (1.25). Therefore, in the frame of the
HAM, we indeed have huge freedom and flexibility. Such kind of freedom and flexibility comes from the homotopy, a basic concept in topology. In theory, this kind of freedom and flexibility provides us great ability to solve some interesting but challenging nonlinear problems (some of them are suggested below in § 1.5 of this chapter), if we can clearly know how to use them in a proper way with stimulated imagination!

In practice, it is suggested to firstly use the zeroth-order deformation equation (1.10), since it works for most of nonlinear problems, as illustrated by Liao [10, 11]. If unsuccessful, one can further attempt a little more complicated zeroth-order deformation equations, such as (1.22). Finally, we emphasize once again that, in theory, one has huge freedom to construct a zeroth-order deformation equation (1.25) satisfying both of (1.26) and (1.27), as long as one clearly knows how to use such kind of freedom.

1.3.2. Spectral HAM and complicated auxiliary operator

Although the HAM provides us great freedom to choose the auxiliary linear operator \( \mathcal{L} \), it might be difficult to solve the linear high-order deformation equation (1.15) or (1.23) exactly, if \( \mathcal{L} \) is complicated. This is mainly because most of linear differential equations have no closed-form solutions, i.e. their solutions are mostly expressed by an infinite series. So, in order to exactly solve high-order deformation equations in the frame of the HAM, we often should choose a reasonable but simple enough auxiliary linear operator \( \mathcal{L} \). This, however, restricts the applications of the HAM.

This is the main reason why only a few simple auxiliary linear operators, such as

\[ \mathcal{L}u = u', \quad \mathcal{L}u = xu' + u, \quad \mathcal{L}u = u' + u, \quad \mathcal{L}u = u'' + u \]

and so on, have been mostly used in the frame of the HAM, where the prime denotes the differentiation with respect to \( x \). These auxiliary linear operators correspond to some fundamental functions such as polynomial, exponential, trigonometric functions and their combination.

There are many special functions governed by linear differential equations. Although many solutions can expressed by these special functions, they are hardly used in the frame of the HAM up to now, because the corresponding high-order deformation equations often become more and more difficult to solve. This is a pity, since in theory the HAM indeed provides us freedom to use special functions to express solutions of many nonlinear differential equations. Currently, Van Gorder [23] made an inter-
esting attempt in this direction. In the frame of the HAM, Van Gorder [23] expressed analytic approximations of the FitzHugh–Nagumo equation by means of error function, Gaussian function and so on. The key is that Van Gorder [23] chose such an auxiliary linear operator
\[ Lu = u'' + \left( \frac{2z^2 - 1}{z} \right) u', \]
where the prime denotes the differentiation with respect to \( z \), and especially such a proper auxiliary function \( H(z) = z|z| \), that the corresponding high-order deformation equations can be solved easily. For details, please refer to Van Gorder [23], Vajravelu and Van Gorder [24] and § 4.6 of this book. This example illustrates once again that the HAM indeed provides us great freedom, i.e. lots of possibilities. The key is how to use such kind of freedom!

Generally speaking, solution of a complicated linear ODE/PDE should be expressed in a series with an infinite number of terms. Mathematically, such a series leads to the larger and larger difficulty to gain higher-order analytic approximations of a nonlinear problem. Fortunately, from physical viewpoint, it is often accurate enough to have analytic approximations with many enough terms. Currently, using the Schmidt-Gram process, Zhao, Lin and Liao [25] suggested an effective truncation technique in the frame of the HAM, which can be used to greatly simplify the right-hand side of the high-order deformation equations, such as (1.15) and (1.23), prior to solving them. In this way, much CPU time can be saved, even without loss of accuracy.

In 2010, Motsa et al. [26, 27] suggested the so-called “spectral homotopy analysis method” (SHAM) using the Chebyshev pseudospectral method to solve the linear high-order deformation equations and choosing the auxiliary linear operator \( L \) in terms of the Chebyshev spectral collocation differentiation matrix [28]. In theory, any a continuous function in a bounded interval can be best approximated by Chebyshev polynomial. So, the SHAM provides us larger freedom to choose the auxiliary linear operator \( L \) and initial guess in the frame of the HAM. It is valuable to expand the SHAM for nonlinear partial differential equations. Besides, it is easy to employ the optimal convergence-control parameter in the frame of the SHAM. Thus, the SHAM has great potential to solve more complicated nonlinear problems, although further modifications in theory and more applications are needed. For the details about the SHAM, please refer to [26, 27] and Chapter 3 of this book.
Chebyshev polynomial is just one of special functions. There are many other special functions such as Hermite polynomial, Legendre polynomial, Airy function, Bessel function, Riemann zeta function, hypergeometric functions, error function, Gaussian function and so on. Since the HAM provides us extremely large freedom to choose auxiliary linear operator $L$ and initial guess, it should be possible to develop a “generalized spectral HAM” which can use proper special functions for some nonlinear problems. Especially, combined the SHAM [26, 27] with the above-mentioned truncation technique suggested by Zhao, Lin and Liao [25], it would be possible to use, when necessary, more complicated auxiliary linear operators in the frame of the HAM so that some difficult nonlinear problems can be solved.

1.3.3. Predictor HAM and multiple solutions

Many nonlinear boundary value problems have multiple solutions. In general, it is difficult to gain these dual solutions by means of numerical techniques, mainly because dual solutions are often strongly dependent upon initial conditions but we do not know how to choose them exactly. Comparatively speaking, it is a little more convenient to use analytic approximation methods to search for multiple solutions of nonlinear problems, since analytic methods admit unknown variables in initial guess.

For example, let us consider a second-order nonlinear differential equation of a two-point boundary value problem:

$$\mathcal{N}[u(x)] = 0, \quad u(0) = a, \quad u(1) = b,$$

where $\mathcal{N}$ is a 2nd-order nonlinear differential operator, $a$ and $b$ are known constants, respectively. Assume that there exist multiple solutions $u(x)$. These multiple solutions must have something different. Without loss of generality, assume that they have different first-order derivative $u'(0) = \sigma$, where $\sigma$ is unknown.

Obviously, different initial guess $u_0(x)$ might lead to multiple solutions. Fortunately, the HAM provides us great freedom to choose initial guess $u_0(x)$. As mentioned before, such kind of freedom is one cornerstone of the HAM. So, in the frame of the HAM, it is convenient for us to choose such an initial guess $u_0(x)$ that it satisfies not only the two boundary conditions $u(0) = a, \ u(1) = b$ but also the additional condition $u'(0) = \sigma$. In this way, the initial guess $u_0(x)$ contains an unknown parameter $\sigma$, called by Liao (see Chapter 8 of [11]) the multiple-solution-control parameter. Then, the analytic approximations gained by the HAM contain at least two un-
known auxiliary parameters: the convergence-control parameter $c_0$ and the multiple-solution-control parameter $\sigma$. As suggested by Liao (see Chapter 8 of [11]), the optimal values of $c_0$ and $\sigma$ can be determined by the minimum of the residual square of governing equations. In this way, multiple solutions of some nonlinear differential equations can be gained, as illustrated by Liao (see Chapter 8 of [11]).

In the frame of the HAM, Abbasbandy and Shivanian [29, 30] developed a differential but rather interesting approach to gain dual solutions, namely the \textit{Predictor HAM} (PHAM). For simplicity, let us use the same equation (1.28) as an example to describe its basic ideas. First of all, an additional condition such as $u'(0) = \sigma$ is introduced with the unknown parameter $\sigma$. Then, in the frame of the HAM, one solves the nonlinear differential equation $N[u(x)] = 0$, but with the two boundary conditions $u'(0) = \sigma$ and $u(1) = b$. Then, $u(0)$, the HAM approximation at $x = 0$, contains at least two unknown parameters: one is the so-called \textit{convergence-control parameter} $c_0$, the other is $\sigma = u'(0)$, called the \textit{multiple-solution-control parameter} by Liao (see Chapter 8 of [11]) in the above-mentioned approach. Substituting the expression of $u(0)$ into the boundary condition $u(0) = a$ gives a nonlinear algebraic equation about $c_0$ and $\sigma$. From the physical viewpoint, $\sigma = u'(0)$ has physical meanings, but the convergence-control parameter $c_0$ does not. If the order of approximation is high enough, one can gain convergent, accurate enough multiple values of $\sigma$ for properly chosen values of $c_0$ in a finite interval, as illustrated in [29, 30]. In this way, one can find multiple solutions of a given nonlinear problem. For details, please refer to Chapter 2 of this book.

In the frame of the HAM, some new branches of solutions for viscous boundary-layer flows were found [31, 32], and the multiple equilibrium-states of resonant waves in deep water [33] and in finite water depth [34] were discovered for the first time, to the best of author’s knowledge. All of these illustrate the potential, novelty and validity of the HAM to give something new and different. This is a superiority of the HAM to numerical methods and some other analytic approximation techniques. Certainly, it is valuable to apply the HAM to discover some new solutions of other nonlinear problems!

1.3.4. \textit{Convergence condition and HAM-based software}

Theoretically speaking, the HAM indeed provides us great freedom to choose initial guess, auxiliary linear operator, convergence-control parame-
ter, equation-type and solution-expression of high-order deformation equa-
tion, and so on. However, it is still not very clear how to use these freedom
in the frame of the HAM, mainly because little mathematical theorems
have been proved in an abstract way.

Some studies on the stability of auxiliary linear operator and
convergence-control parameter of the HAM are described in Chapter 4 of
this book. Some current works about convergence condition of the HAM
series are described in Chapter 5.

It should be especially emphasized that Park and Kim [35, 36] success-
fully applied the HAM to solve a few classic problems in finance, and gave
convergence conditions for their analytic approximations. It is rather inter-
esting that they even gave an error estimation for their analytic approxima-
tions in [36]. Currently, Park and Kim used the HAM to solve an abstract
linear problem with respect to bounded linear operators from a Banach
space to a Banach space, and rigorously proved that the homotopy solution
exists in the sense that a series of the problem converges in a Banach norm
sense if the linear operator satisfies some mild conditions. Their fantastic
works are very important, and might pioneer a new research direction and
style (i.e. abstract proof) in the frame of the HAM. Such kind of abstract
mathematical theorems in the frame of the HAM are more valuable and
useful, if nonlinear governing equations and especially the influence of the
convergence-control parameter on the convergence could be considered.

On the other side, the HAM has been successfully applied to numerous
nonlinear problems in various fields of science and engineering. These ap-
plications show the general validity and novelty of the HAM. Unfortunately,
it is impossible to mention all of them here in details. As examples among
these numerous applications, a HAM-based approach about boundary-layer
flows of nanofluid is given in Chapter 6 of this book. In addition, an appli-
cation of the HAM for time-fractional boundary-value problem is illustrated
in Chapter 7.

To simplify some applications of the HAM, two HAM-based software
were developed. The HAM-based Maple package NOPH (version 1.0.2)
for periodic oscillations and limit cycles of nonlinear dynamic systems is
described in Chapter 8 of this book with various applications. It is free
available online at

http://numericaltank.sjtu.edu.cn/NOPH.htm

with a simple user’s guide. Besides, the HAM-based Mathematica package
BVPh (version 2.0) for coupled nonlinear ordinary differential equations
with boundary conditions at multiple points are given in Chapter 9 of this book. It is free available online at

http://numericaltank.sjtu.edu.cn/BVPh.htm

with a simple user’s guide and some examples of application. Both of these two HAM-based software are easy-to-use and user-friendly. They greatly simplify some applications of the HAM, and are especially helpful for the beginners of the HAM.

1.4. Relationships to other methods

In pure mathematics, the homotopy is a fundamental concept in topology and differential geometry. The concept of homotopy can be traced back to Jules Henri Poincaré (1854–1912), a French mathematician. A homotopy describes a kind of continuous variation (or deformation) in mathematics. For example, a circle can be continuously deformed into a square or an ellipse, the shape of a coffee cup can deform continuously into the shape of a doughnut but cannot be distorted continuously into the shape of a football. Essentially, a homotopy defines a connection between different things in mathematics, which contain same characteristics in some aspects. In pure mathematics, the homotopy is widely used to investigate existence and uniqueness of solutions of some equations, and so on.

In applied mathematics, the concept of homotopy was used long ago to develop some numerical techniques for nonlinear algebraic equations. The so-called “differential arc length homotopy continuation method” were proposed in 1970s by Keller [37, 38]. However, the global homotopy methods can be traced as far back as the work of Lahaye [39] in 1934. To solve a nonlinear algebraic equation \( f(x) = 0 \) by means of the homotopy continuation method, one first constructs such a homotopy

\[
H(x, q) = q f(x) + (1 - q) g(x), \tag{1.29}
\]

where \( q \in [0, 1] \) is the homotopy parameter, \( g(x) \) is a function for which a zero is known or readily obtained. As discussed by Wayburn and Seader [40], the choice of \( g(x) \) is arbitrary, but the two most widely used functions are the Newton homotopy

\[
H(x, q) = q f(x) + (1 - q) \left[ f(x) - f(x_0) \right],
\]

and the fixed-point homotopy

\[
H(x, q) = q f(x) + (1 - q) (x - x_0),
\]
where \( x_0 \) is an arbitrary starting point. The locus of solutions defines the homotopy path that is tracked with some continuation method. Consequently, homotopy continuation methods consist not only of the homotopy equation itself, but also the homotopy path tracking method, i.e. of some continuation strategy.

Homotopy continuation methods are usually based upon differentiating the homotopy equation (1.29) with respect to the arc length \( s \), which gives the equation

\[
\frac{\partial H}{\partial x} \frac{dx}{ds} + \frac{\partial H}{\partial q} \frac{dq}{ds} = 0.
\]

Taking into account the arc-length relation

\[
\left( \frac{dx}{ds} \right)^2 + \left( \frac{dq}{ds} \right)^2 = 1
\]

and the initial condition \( H(x_0, 0) = 0 \), we obtain an initial value problem. Then, path tracking based on the initial value problem is numerically carried out with a predictor-corrector algorithm to gain a solution of the original equation \( f(x) = 0 \). Some elegant theorems of convergence are proved in the frame of the homotopy continuation method. For details of the homotopy continuation method, please refer to [41–49].

Unlike the homotopy continuation method that is a global convergent numerical method mainly for nonlinear algebraic equations, the HAM is a kind of analytic approximation method mainly for nonlinear differential equations. So, the HAM is essentially different from the homotopy continuation method, although both of them are based on the homotopy, the basic concept of the topology. Note that the HAM uses much more complicated homotopy equation (1.10) or (1.22) than (1.29) for the homotopy continuation method. Furthermore, the HAM provides larger freedom to choose the auxiliary linear operator \( \mathcal{L} \). Most importantly, the so-called convergence-control parameter \( c_0 \) is introduced for the first time, to the best of our knowledge, in the homotopy equation (1.10) or (1.22) so that the HAM provides us a convenient way to guarantee the convergence of series series. Note that the homotopy equation (1.29) of the homotopy continuation method does not contain such kind of convergence-control parameter at all. So, the convergence-control parameter \( c_0 \) is indeed a novel. In fact, it is the convergence-control parameter \( c_0 \) which differs the HAM from all other analytic approximation methods.

In addition, the HAM logically contains many other analytic approximation methods and thus is rather general. For example, it has been generally
proved [10–13] that the Lyapunov’s artificial small parameter method [5],
the Adomian decomposition method [6–8], the δ-expansion method [18] are
only special cases of the HAM for some specially chosen auxiliary linear
operator \( L \) and convergence-control parameter \( c_0 \). Furthermore, the so-
called “optimal homotopy asymptotic method” [22] developed in 2008 is
also a special case of the homotopy equation (1.22) when \( K \to +\infty \), too,
as pointed out by Liao (see § 6.3 of Liao’s book [11]).

Especially, the so-called “homotopy perturbation method” (HPM) [19]
proposed by Jihuan He in 1998 (six years later after Liao [9] proposed the
early HAM in 1992) was only a special case of the HAM when \( c_0 = -1 \), as
proved in [20], and thus it has “nothing new except its name” [20]. Some
results given by the HPM are divergent even in the whole interval except
the given initial/boundary conditions, and thus “it is very important to
investigate the convergence of approximation series, otherwise one might
get useless results”, as pointed out by Liang and Jeffrey [21]. For more

In addition, even the famous Euler transform in calculus can be derived
in the frame of the HAM (see Chapter 5 of Liao’s book [11]). This provides
us a theoretical cornerstone for the validity and generality of the HAM.

In summary, based on the concept of homotopy topology, the HAM
is a novel analytic approximation method for highly nonlinear problems,
with great freedom and flexibility to choose equation-type and solution ex-
pression of high-order approximation equations and also with a convenient
way to guarantee the convergence, so that it might overcome restrictions of
perturbation techniques and other non-perturbation methods.

1.5. Chance and challenge: some suggested problems

Any truly new methods should give something novel and/or different, or
solve some difficult problems that can not be solved with satisfaction by
other methods.

Unlike other analytic approximation methods, the HAM provides us
great freedom and flexibility to choose equation-type and solution expres-
sion of high-order approximation equations, and especially a simple way
to guarantee the convergence of solution series. Thus, the HAM provides
us a large possibility and chance to give something novel or different, and
to attack some difficult nonlinear problems. For example, some new so-
lutions [31, 32] of boundary-layer flows have been found by means of the
HAM, which had been neglected even by numerical techniques and had
been never reported. Some analytic approximations for the optimal exercise boundary of American put option were given, which are valid from a couple of years (see [50, 51]) up to even 20 years (see Chapter 13 of Liao’s book [11]) prior to expiry, and thus much better than the asymptotic/perturbation approximations that are often valid only for a couple of days or weeks. Besides, the HAM has been successfully employed to solve some complicated nonlinear PDEs: the multiple equilibrium-states of resonant waves in deep water [33] and in finite water depth [34] were discovered by means of the HAM for the first time, to the best of our knowledge, which greatly deepen and enrich our understandings about resonant waves.

All of these successful applications show the originality, validity and generality of the HAM for nonlinear problems, and encourage us to apply the HAM to attack some famous, challenging nonlinear problems. Some of these problems are suggested below for the HAM community, especially for brave, enterprising, young researchers.

1.5.1. Periodic solutions of chaotic dynamic systems

It is well known that chaotic dynamic systems have the so-called “butterfly effect” [52, 53], say, the computer-generated numerical simulations have sensitive dependence to initial conditions (SDIC). For example, the nonlinear dynamic system of Lorenz equations [52]

\[
\begin{align*}
\dot{x} &= \sigma(y - x), \\
\dot{y} &= r x - y - x z, \\
\dot{z} &= x y - b z,
\end{align*}
\]

has chaotic solution in case of \( r = 28, b = 8/3 \text{ and } \sigma = 10 \) for most of given initial conditions \( x_0, y_0, z_0 \) of \( x, y, z \) at \( t = 0 \). However, for some special initial conditions such as

\[
\begin{align*}
x_0 &= -13.7636106821, \quad y_0 = -19.5787519424, \quad z_0 = 27; \\
x_0 &= -9.1667531454, \quad y_0 = -9.9743951128, \quad z_0 = 27; \\
x_0 &= -13.5683173175, \quad y_0 = -19.1345751139, \quad z_0 = 27,
\end{align*}
\]

the above dynamic system of Lorenz equation has unstable periodic solutions, as reported by Viswanath [54].

A periodic solution \( u(t) \) with the period \( T \) has the property

\[ u(t) = u(t + nT) \]
for arbitrary time \( t \geq 0 \) and arbitrary integers \( n \), even if \( t \to +\infty \) and \( n \to \infty \). This property cannot be checked strictly by means of numerical approaches, since all numerical integration simulations are gained in a finite interval of time. Naturally, a periodic solution should be expressed analytically by periodic base functions such as trigonometric functions. So, theoretically speaking, it is inherently better to use analytic approximation methods to search for periodic solutions of chaotic dynamic systems than numerical ones.

In fact, as illustrated by Liao in Chapter 13 of his book [10], the HAM can be employed to gain periodic solution of nonlinear dynamic systems. Can we employ the HAM to gain the above-mentioned unstable periodic solutions of Lorenz equation found by Viswanath [54]? More importantly, it would be very interesting if the HAM could be employed to find some new periodic solutions of Lorenz equation with physical parameters leading to chaos! This is mainly because Lorenz equation is one of the most famous ones in nonlinear dynamics and nonlinear science.

### 1.5.2. Periodic orbits of Newtonian three-body problem

Let us consider one of the most famous problem in mechanics and applied mathematics: the Newtonian three-body problem, say, the motion of three celestial bodies under their mutual gravitational attraction. Let \( x_1, x_2, x_3 \) denote the three orthogonal axes. The position vector of the \( i \)th body is expressed by \( \mathbf{r}_i = (x_{1,i}, x_{2,i}, x_{3,i}) \), where \( i = 1, 2, 3 \). Let \( T \) and \( L \) denote the characteristic time and length scales, and \( m_i \) the mass of the \( i \)th body, respectively. Using Newtonian gravitation law, the motion of the three bodies are governed by the corresponding non-dimensional equations

\[
\ddot{x}_{k,i} = \sum_{j=1, j \neq i}^{3} \rho_j \frac{(x_{k,j} - x_{k,i})}{R_{i,j}^3}, \quad k = 1, 2, 3, \tag{1.34}
\]

where

\[
R_{i,j} = \left[ \sum_{k=1}^{3} (x_{k,j} - x_{k,i})^2 \right]^{1/2} \tag{1.35}
\]

and

\[
\rho_i = \frac{m_i}{m_1}, \quad i = 1, 2, 3 \tag{1.36}
\]

denotes the ratio of the mass.
According to H. Poincaré, orbits of three-body problem are unintegrable in general. Although chaotic orbits of three-body problems widely exist, three families of periodic orbits were found:

(1) the Lagrange–Euler family, dating back to the analytical solutions in the 18th century (one recent orbit was given by Moore [55]);
(2) the Broucke–Hadjidemetriou–Hénon family, dating back to the mid-1970s [56–61];
(3) the Figure-8 family, discovered in 1993 by Moore [55] and extended to the rotating cases [62–65].

Note that nearly all of these reported periodic orbits are planar. In 2013, Šuvakov and Dmitrašinović [66] found that there exist four classes of planar periodic orbits of three-body problem, with the above three families belonging to one class. Besides, they reported three new classes of planar periodic orbits and gave the corresponding initial conditions for each class. For the details of their 15 planar periodic orbits, please refer to the gallery [67].

Šuvakov and Dmitrašinović [66] found these new classes of planar periodic orbits by means of an iterative numerical integration approach without using multiple precision. So, it is unknown whether or not the numerical simulations depart the corresponding periodic orbits for rather large time, i.e. \( t \to \infty \). As mentioned before, it is better and more natural to express a periodic solution \( u(t) \) with the period \( T \) in series of periodic base functions (with the same period \( T \)) so that \( u(t) = u(t + nT) \) can hold for arbitrary integer \( n \) and arbitrary time \( t \) even if \( t \to \infty \). Thus, it is valuable to apply the HAM to double check all of the reported periodic orbits in [66], and more importantly, to find some completely new periodic orbits!

Note that nearly all of the periodic orbits of three-body problem reported up to now are planar. Therefore, it is valuable and interesting if the HAM can be applied to find some periodic orbits of Newtonian three-body problems, which are not planar, i.e. three dimensional. Mathematically speaking, we should determine such unknown initial positions \( r_1, r_2, r_3 \), unknown initial velocities \( \dot{r}_1, \dot{r}_2, \dot{r}_3 \) and unknown corresponding mass-ratios \( \rho_1, \rho_2, \rho_3 \) of three bodies in the frame of the HAM that Eqs. (1.34) have periodic solution \( x_{k,i}(t) = x_{k,i}(t + nT) \) for arbitrary time \( t \) and integer \( n \), where \( T \) is the unknown corresponding period to be determined, and \( i, k = 1, 2, 3 \). This is a valuable, interesting but challenging problem for the HAM community.
1.5.3. **Viscous flow past a sphere**

One of the most famous, classical problems in fluid mechanics is the steady-state viscous flow past a sphere [68–75], governed by the Navier-Stokes equation, i.e. a system of nonlinear partial differential equations. Consider the steady-state viscous flow past a sphere in a uniform stream. How large is the drag of the sphere due to the viscosity of fluid?

To study the steady-state viscous flow past a sphere, the spherical coordinates $\mathbf{r} = (r, \theta, \phi)$ is often used. Since the problem has axial symmetry, one can use the Stokes stream function $\psi(r, \theta)$ defined through the following relations:

$$v_r = \frac{1}{r^2 \sin(\theta)} \psi_\theta, \quad v_\theta = -\frac{1}{r \sin(\theta)} \psi_r, \quad v_\phi = 0. \quad (1.37)$$

The stream function $\psi(r, \theta)$ is governed by the dimensionless equation

$$D^4 \psi = \frac{R}{r^2} \left[ \frac{\partial (\psi, D^2 \psi)}{\partial (r, \mu)} + 2D^2 \psi L \psi \right], \quad (1.38)$$

subject to the boundary conditions

$$\psi(1, \mu) = 0, \quad (1.38a)$$

$$\left. \frac{\partial \psi(r, \mu)}{\partial r} \right|_{r=1} = 0, \quad (1.38b)$$

$$\lim_{r \to \infty} \frac{\psi(r, \mu)}{r^2} = \frac{1}{2} (1 - \mu^2), \quad (1.38c)$$

where $R = aU_\infty/\nu$ is the Reynolds number and

$$\mu \equiv \cos(\theta), \quad (1.39)$$

$$D^2 \equiv \frac{\partial^2}{\partial r^2} + \frac{1 - \mu^2}{r^2} \frac{\partial^2}{\partial \mu^2}, \quad (1.40)$$

$$L \equiv \frac{\mu}{1 - \mu^2} \frac{\partial}{\partial r} + \frac{1}{r} \frac{\partial}{\partial \mu}. \quad (1.41)$$

Here, $a$ denotes the radius of the sphere and $U_\infty$ the uniform stream velocity at infinity, respectively, according to the notation of Proudman and Pearson [71]. As mentioned by Liao [74], the drag coefficient reads

$$C_D = \frac{4}{R} \int_{-1}^{1} \left. \left( -p \mu + \frac{\partial^2 \psi}{\partial r^2} \right) \right|_{r=1} d\mu, \quad (1.42)$$

where the pressure $p$ is given by

$$p = -\int_{\mu}^{1} \frac{1}{(1 - \mu^2)} \left. \frac{\partial^3 \psi}{\partial r^3} \right|_{r=1} d\mu. \quad (1.43)$$
Unfortunately, neither the linearization method [68–70] nor the perturbation techniques [71, 72] can provide an analytic approximation of drag coefficient $C_D$ valid for $R_d > 3$, where $R_d = d U_\infty / \nu = 2R$ for the diameter $d$ of the sphere. Especially, the 3rd-order multiple-scale perturbation approximation of $C_D$ given by Chester and Breach [72] was valid even in a smaller interval of Reynolds number than the 2nd-order multiple-scale perturbation result of Proudman and Pearson [71]. This implies the invalidity of perturbation methods for this famous problem. So, “the idea of using creeping flow to expand into the high Reynolds number region has not been successful”, as pointed out by White in his textbook [76]. Besides, the method of renormalization group can not essentially modify these analytic results [75], either.

In 2002, Liao [74] employed the HAM to solve the steady-state viscous flow past a sphere and gained a analytic approximation of drag coefficient $C_D$, which agree well with experimental data in a considerably larger interval $R_d \leq 30$. However, the corresponding experiments indicate that the steady-state viscous flow past a sphere exists until $R_d \approx 100$. So, strictly speaking, this HAM result given in [74] is not satisfactory.

Theoretically speaking, it is very interesting and valuable if one can give an accurate enough analytic result of the drag coefficient $C_D$ valid for the steady-state viscous flow past a sphere up to $R_d \approx 100$, mainly because it is one of the most famous, classical problems in fluid mechanics with a history of more than 150 year!

Can we solve this famous, classical problem by means of the HAM?

1.5.4. Viscous flow past a cylinder

The steady-state viscous flow past an infinite cylinder is also one of the most famous, classical problems in fluid mechanics with a long history. For the steady-state viscous flow past an infinite cylinder, it is natural to use cylindrical coordinates $\vec{r} = (r, \theta, z)$. Since the problem is two dimensional, it is convenient to use the Lagrangian stream function $\psi(r, \theta)$ defined by Proudman and Pearson [71]:

$$u_r = \frac{1}{r} \frac{\partial \psi}{\partial \theta}, \quad u_\theta = -\frac{\partial \psi}{\partial r}, \quad u_z = 0.$$  \hspace{1cm} (1.44)

The stream function $\psi(r, \theta)$ is governed by

$$\nabla^2_r \psi(r, \theta) = -\frac{R}{r} \frac{\partial (\psi \cdot \nabla^2_r \psi)}{\partial (r, \theta)}.$$  \hspace{1cm} (1.45)
subject to the boundary conditions

\[ \psi(r = 1, \theta) = 0, \]  
\[ \frac{\partial \psi(r, \theta)}{\partial r} \bigg|_{r=1} = 0, \]  
\[ \lim_{r \to \infty} \frac{\psi(r, \theta)}{r} = \sin(\theta), \]

where

\[ \nabla_r^2 \equiv \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2}, \quad \nabla_r^4 \equiv \nabla_r^2 \nabla_r^2. \]

Here, \( R = aU_\infty/\nu \) is the Reynolds number, \( a \) and \( U_\infty \) denote the radius of cylinder and the uniform stream velocity at infinity, respectively.

As reviewed in [75], neither the linearization method nor perturbation technique can give good analytic approximation of the drag coefficient \( C_D \) of a cylinder for \( R \geq 3 \). In fact, even the method of renormalization group cannot modify these results greatly [75]. So, it is still an open question.

Theoretically speaking, it is valuable to gain an accurate analytic expression of drag coefficient \( C_D \) valid for large Reynolds number up to \( R \approx 40 \), beyond which the periodic Von Kármán vortex occurs. This is mainly because it is one of the most famous, historical problem in fluid mechanics.

Can this famous, classical problem be solved by means of the HAM?

1.5.5. Nonlinear water waves

The HAM has been successfully applied to solve some nonlinear wave equations. Especially, in the frame of the HAM, the multiple equilibrium-states of resonant waves in deep water [33] and in finite water depth [34] were discovered for the first time, to the best of the author’s knowledge. Thus, the HAM provides us a convenient tool to investigate some complicated wave problems.

Strictly speaking, water waves are governed by Euler equation with two nonlinear boundary conditions satisfied on an unknown free surface, which however are rather difficult to solve in general. Based on the exact Euler equation, some simplified wave models for shallow water waves, such as the KdV equation [77], Boussinesq equation [78], Camassa–Holm (CH) equation [79], and so on, are derived by assuming the existence of some small physical parameters in shallow water. Although these shallow water wave equations are much simpler than the exact Euler equation, they can well ex-
plain many physical phenomena, such as soliton waves, wave propagations and interactions in shallow water, wave breaking, and so on.

For example, the celebrated Camassa–Holm (CH) equation [79]

\[ u_t + 2\omega u_x - u_{xxt} + 3uu_x = 2u_xu_{xx} + uu_{xxx}, \tag{1.46} \]

subject to the boundary condition

\[ u = 0, \quad u_x = 0, \quad u_{xx} = 0, \quad \text{as} \quad x \to \pm \infty, \tag{1.47} \]

can model both phenomena of soliton interaction and wave breaking (see [80]), where \( u(x,t) \) denotes the wave elevation, \( x, t \) are the temporal and spatial variables, \( \omega \) is a constant related to the critical shallow water wave speed, the subscript denotes the partial differentiation, respectively. Mathematically, the CH equation is integrable and bi-Hamiltonian, thus possesses an infinite number of conservation laws in involution [79]. In addition, it is associated with the geodesic flow on the infinite dimensional Hilbert manifold of diffeomorphisms of line (see [80]). Thus, the CH equation (1.46) has many intriguing physical and mathematical properties. As pointed out by Fushsteiner [81], the CH equation (1.46) even “has the potential to become the new master equation for shallow water wave theory”.

Especially, when \( \omega = 0 \), the CH equation (1.46) has the peaked solitary wave

\[ u(x,t) = c \exp(-|x-ct|), \]

which was found first by Camassa and Holm [79]. The first derivative of the peaked solitary wave is discontinuous at the crest \( x = ct \). Like the CH equation, many shallow water equations admit peaked and/or cusped solitary waves. These equations with peaked and/or cusped solitary waves have been widely investigated mathematically, and thousands of related articles have been published. However, to the best of the author’s knowledge, peaked and cusped solitary waves have never been gained directly from the exact Euler equation! This is very strange. Logically speaking, since these simplified equations (like the CH equation) are good enough approximations of the Euler equation in shallow water, the exact Euler equation should also admit the peaked and/or cusped solitary waves as well.

Can we gain such kind of peaked and/or cusped solitary waves of the exact wave equation by means of the HAM, if they indeed exist? Either positive or negative answers to this question have important scientific meanings. If such kind of peaked solutions of the exact wave equation indeed exist, it can greatly enrich and deepen our understandings about peaked
solitary waves. If the peaked solitary waves given by the exact wave equation exists mathematically but is impossible in physics, we had to check the physical validity of the peaked solitary waves. So, this is an interesting and valuable work, although with great challenge. For some attempts in this direction, please refer to Liao [82], who proposed a generalized wave model based on the symmetry and the fully nonlinear wave equations, which admits not only the traditional waves with smooth crest but also peaked solitary waves. It is found that the peaked solitary waves satisfy Kelvin’s theorem everywhere. Besides, these peaked solitary waves include the famous peaked solitary waves of the Camassa–Holm equation. So, the generalized wave model [82] is consistent with the traditional wave theories. It is found [82] that the peaked solitary waves have some unusual characteristics quite different from the traditional ones, although it is still an open question whether or not they are reasonable in physics if the viscosity of fluid and the surface tension are considered.

In addition, the so-called “rogue wave” [83, 84] is a hot topic of nonlinear waves. Certainly, it is valuable to apply the HAM to do some investigations in this field.

In summary, it is true that the problems suggested above are indeed difficult, but very valuable and interesting in theory. In fact, there are many such kind of interesting but difficult problems in science, engineering and finance. It should be emphasized that, unlike all other analytic approximation methods, the HAM provides us great freedom and flexibility to choose equation-type and solution expression of high-order approximation equations, and besides a convenient way to guarantee the convergence of solution series. As pointed out by Georg Cantor (1845–1918), “the essence of mathematics lies entirely in its freedom”. Hopefully, the great freedom and flexibility of the HAM might create some novel ideas and inspire some brave, enterprising, young researchers with stimulated imagination to attack them with satisfactory, much better results.

Chance always stays with challenges!

Acknowledgment

This work is partly supported by National Natural Science Foundation of China (Approval No. 11272209), the Foundation for Shanghai Leading Scientists, and State Key Laboratory of Ocean Engineering (Approval No. GKZD010056).
References

2301 (2008).


[37] H.B. Keller, Numerical solution of bifurcation and nonlinear eigenvalue prob-


[67] http://suki.ipb.ac.rs/3body/.
[78] J. Boussinesq, Théorie des ondes et des remous qui se propagent le long d'un
canal rectangulaire horizontal, en communiquant au liquide contenu dans ce
canal des vitesses sensiblement pareilles de la surface au fond, *Journal de

[79] R. Camassa and D.D. Holm, An integrable shallow water equation with

[80] A. Constantin, Existence of permanent and breaking waves for a shallow

[81] B. Fuchssteiner, Some tricks from the symmetry-toolbox for nonlinear equa-
tions: generalizations of the Camassa-Holm equation, *Physica D*. 95:

2012.

[83] I. Didenkulova and E. Pelinovsky, Rogue waves in nonlinear hyperbolic sys-

[84] A. Calini and C. M. Schober, Dynamical criteria for rogue waves in nonlinear