On the Homoclinic Tangles of Henri Poincaré

Qiudong Wang

Author’s note: The purpose of this essay is to put my recent work with Ali Oksasoglu, discussed in Section 5, in historic perspective. Dynamical systems is a vast research subject and this short essay is focused only on one issue. Even within the scope of this one issue, this essay is not intended as a survey of history, though most things I chose to discuss in the first four sections are of historic implication. Needless to say that my choices are subjected to the bias of my own academic taste.

1. King Oscar II’s prize on the N-body problem

Our story started with a prize established by the King Oscar II of Sweden and Norway in 1888 for solving the Newtonian N-body problem. The idea of this prize was from Gösta Mittag-Leffler, his Majesty’s science adviser. The prize committee was formed of Gösta Mittag-Leffler, Karl Weierstrass, and Charles Hermite, and the first task of the committee was to define precisely the meaning of “solving the N-body problem”. This task was delegated to Weierstrass, and the formulation he came up with, formally introduced in the official announcement for the prize in Acta Mathematica, vol. 7, of 1885-1886, was as follows

Given a system of arbitrarily many mass points that attract each other according to Newton’s law, under the assumption that no two points ever collide, try to find a representation of the coordinates of each point as a series in a variable that is some known function of time and for all of whose values the series converges uniformly.

The proposed power series solutions were later constructed by Karl Sundman for the 3-body problem ([35]) and by myself for all \(N > 3\) ([37]). In both cases, however, the power series constructed converged so slowly that they were practically useless. The method of construction I provided, fulfilling all requirements of the question as proposed, was in fact tricky but surprisingly simple. A question with a tricky, simple and useless answer should not be the kind the prize committee ought to ask. This, therefore, was the first mathematical mistake made by the committee in connection with King Oscar’s prize.

Let us now move on with the story. Henri Poincaré, the greatest mathematical genius of the time, started working on the proposed problem. In a letter addressed to Mittag-Leffler, he claimed that he has proved a stability result for the restricted three-body problem. He wrote ([13], page 44)

In this particular case, I have found a rigorous proof of stability and a method of placing precise limits on the elements of the third body... I now hope that I will be able to attack the general case and ... if not completely resolve the problem (of this I have little hope), then at least found sufficiently complete results to send into the competition.
A paper was then submitted, and the prize was awarded to him. His paper was dually refereed by Mittag-Leffler and Weierstrass, and the latter asserted in his report to the former that ([13], page 44)

*I have no difficulty in declaring that the memoir in question deserves the prize. You may tell your Sovereign that this work can not, in truth, be considered as supplying a complete solution to the question we proposed, but it is nevertheless of such importance that its publication will open a new era in the history of celestial mechanics. His Majesty’s goal in opening the contest can therefore be considered attained.*

It was, however, soon realized that Poincaré’s prize winning paper contained a fatal mathematical error and the stability result he claimed to Mittag-Leffler was wrong. It appeared that Poincaré himself was the one who spotted the fatal error in the middle of trying to answer certain questions raised by Edvard Phragman, the editor of Acta Mathematica responsible for copyediting his paper for publication.

The prevailing opinion in later times concerning the mistake and the award was perhaps reflected best in the following passage, which was written by Forest Ray Moulton in 1912 ([23])

*While the error was unfortunate, there is not the slightest doubt that in spite of it, and even had it been known at the time, the prize was correctly bestowed. If all the parts affected by the error are omitted, the memoir still remains one whose equal in originality, in results secured, and in extent of valuable field opened, is difficult to find elsewhere. There are but few men, even of high reputation, who have produced more in their whole lives that was really new and valuable than that which was correct in the original investigation submitted by Poincaré.*

It would be unwise to argue with Professor Moulton against Poincaré’s originality and the greatness of his mathematical achievement. But mistakes were mistakes, and the mistake made by the committee to award the prize to a paper with a fatal mathematical error was not entirely excusable. Besides, it actually cost Poincaré real money. At the time the mistake was uncovered, his original submission was already in print. Mittag-Leffler decided to make a recall and he ordered all recalled copies with Poincaré’s original paper physically destroyed. Poincaré paid twice as much as the prize money he received to cover the cost of the recall. Anyway, Mittag-Leffler was in a real difficult situation, and it appeared that he walked a thin line in trying to maintain a reasonable ethic standard and at the same time control a potential fallout.

In trying to understand the mathematical consequences of his mistake, Poincaré discovered the homoclinic tangle therefore gave birth to the chaos theory, a mathematical theory of great influence in later times. Poincaré’s paper published by Acta Mathematica ([28]) after the recall is a true masterpiece and history has proved that it equaled all that was asserted in Weierstrass’s report.

### 2. Discovery of Homoclinic Tangles

To understand the mathematical error Poincaré made in his paper and the discovery that followed, we need to start with the geometric point of view Poincaré introduced in the study of ordinary differential equations ([27]). Let \(x = (x_1, \cdots, x_n)\) be an
$n$-vector and $t$ be the time. Let

$$\frac{dx}{dt} = f(x)$$

be a set of ordinary differential equations for $x$, and $x(t, x(0))$ be the solution of (2.1) satisfying $x(0, x(0)) = x(0)$. Before the time of Poincaré, mathematicians had tried to solve equation (2.1) by deriving analytic formula for $x(t, x(0))$. Since solutions in closed form are in general not attainable, power series were used as a substitute. At that time, there were very practical and rather urgent needs in Astronomy and in Navigation to acquire accurate solutions of the $N$-body problem. Nowadays, modern computer can do the job required through numerical integration; but at the time people had to rely on power series solution and it was not uncommon for a mathematician to spend his entire career computing power series solution of the 3-body problem, using all the tricks one could possibly imagine in attaining more accurate solution that can better predict the position of, say, the Moon. However, power series only converge on finite time interval, and the size of the convergence interval varies depending on the location of the solution. Weierstrass’s formulation of the King’s prize problem was a reflection of the mainstream perception of his time. He equated the task of solving the $N$-body problem to the task of finding a power series solution that converges for all time.

Poincaré’s geometric point of view, which is by the way simple, is to view an $n$-vector as a point in $\mathbb{R}^n$, which he called the phase space and to view solutions as a collection of non-intersecting curves in phase space. He pointed out that, as a mathematical problem, the questions we should ask in the study of equation (2.1) are (a) what kind of solution curves are allowed by equation (2.1), and (b) in what way all these solution curves fit together.

The ordinary differential equation Poincaré was toiling on for the King’s prize was not exactly like equation (2.1). It is in the form of

$$\frac{dx}{dt} = f(x) + \varepsilon g(x, t)$$

where $\varepsilon$ is a small parameter and $g(x, t)$ is a function that is periodic in $t$ of, say period $T$. One good thing about the restricted 3-body problem, which was the subject of Poincaré’s investigation, is that the equation for $\varepsilon = 0$ is completely solvable so he knew everything about the solution curves and the way they fit together. In particular, he knew that there is an equilibrium solution, to which a family of solutions asymptotically approach in both the backward and the forward time. These solutions Poincaré called homoclinic solutions. Homoclinic solutions form a nice invariant surface in phase space. Now the question is what happens to this nice invariant surface when $\varepsilon$ is not zero. Here Poincaré overlooked the possibility that periodic perturbation could break the invariant surface, and he used implicitly an unbroken surface in the case of $\varepsilon \neq 0$ in deriving his stability conclusion.

Poincaré explained the situation later by using equations with a two dimensional phase space ([29]). To study the solutions of (2.2), Poincaré explained, is the same as to study the iterations of a map that is defined by the solutions of equation (2.2). Let $x(t, x(0))$ be the solution of (2.2) satisfying $x(0, x(0)) = x(0)$, the map Poincaré iterated is $x(0) \rightarrow x(T, x(0))$, which is defined from $\mathbb{R}^2$ to $\mathbb{R}^2$. This map is the $time-T$
map. The trajectory of a homoclinic solution in phase space is an invariant loop for the time-T map of the unperturbed equation. Small perturbation, unfortunately, would usually break the loop into two intersecting curves, as is shown in Figure 1(a).

Figure 1. Homoclinic intersection and Homoclinic tangle

What Poincaré discovered was when the loop is broken into two intersecting curves, which he knew by then as inevitable, they would be forced to intersect each other to form a web, and the structure of this web appeared to be incomprehensibly complicated. See Figure 1(b). For solutions in this complicated mesh, which he called as a homoclinic tangle, there would be no well-defined destination, therefore no possibility for dynamical stability.

3. Smale’s Horseshoe

From the time of Henri Poincaré to early 1960’s, many people, including Birkhoff ([7]), Cartwright and Littlewood ([9]), Levinson ([15]), Sitnikov ([34]) and Alekseev ([1]), had studied more systems of differential equation from celestial mechanics, material science and electric circuits, and they had proved in quite a few of these equations that homoclinic tangle exists. Some had also come to the conclusion that periodic solutions accumulate in homoclinic tangle. Birkhoff and Levinson even used symbolic sequence to code the solutions.

Steven Smale studied homoclinic tangle in early 1960’s and he concluded that there is a simple geometric structure embedded in all homoclinic tangle, and this structure produces accumulation of periodic solutions and other complicated dynamical behavior ([32]). The geometric structure Smale introduced is as follows. Let us start with a 2D square. We first compress it in vertical direction and stretch it in horizontal direction to make a thin and long strip. We then fold the strip and put it back on the original square. See Figure 2(a). This defines a map, which Smale called a horseshoe map because the image of the map looks like one.

Under the horseshoe map, part of the square is mapped out and part is mapped back into the square. Smale observed that, there is a subset that would stay inside the square forever under the forward and the backward iterations of the horseshoe map, and this set has a complicated but thoroughly understandable structure. A conceptual way to comprehend the structure of this invariant set is to forever replace
every square by four smaller squares inside, starting with the original square. See Figure 2(b). It is not hard to show that this invariant subset contains infinitely many unstable periodic orbits for the horseshoe map.

![Figure 2. Smale’s horseshoe](image)

Smale illustrated that every homoclinic tangle contains a horseshoe. This is shown in Figure 3. Starting with a horizontal strip containing a segment of the stable curve, as drew in Figure 3, the image of the strip will eventually follow a segment of the unstable curve to form a horseshoe.

![Figure 3. Horseshoe embedded inside of a homoclinic tangle](image)

The elegance and the simplicity of Smale’s construction made a very complicated mathematical situation accessible to even non-mathematicians. Together with the later discoveries of similar elegance and simplicity, such as the Lorenz butterfly ([19]), Li-York’s *period three implies chaos* ([18]), and Feigenbaum’s *periodic doubling diagram* ([11]), it generated great enthusiasm for the chaos theory in the general scientific community and carried the banner of the theory of chaos and modern dynamical systems wide and far into many other scientific disciplines.

4. HENON MAPS AND HOMOCLINIC TANGENCY

Up to now, nobody has proved that the N-body problem is unsolvable. However, we have learnt from Henri Poincaré that one pre-requisite for solving the N-body problem is to understand the dynamics of certain homoclinic tangle. Smale’s horseshoe, we should caution, is only a *participating part* in homoclinic tangles.

To study homoclinic tangles, on the other hand, we do not have to start with a differential equation. Maps with homoclinic tangle are easy to come by and iterating
a map is much easier than solving a differential equation. Therefore it was almost like a liberation when the attention was shifted gradually from equation to map. Questions of independent mathematical interest arose and dynamical systems as a research subject exploded. In this history of fascinating progress, starting from early 1960’s, two maps have commanded tremendous attention and they are the Anosov diffeomorphisms ([3]) and the Henon maps ([14]).

Anosov diffeomorphism provided a penetrating point and a base for mathematician to forge new concept and to develop new technique to study homoclinic tangles. The study of these uniformly hyperbolic systems has been critical in shaping the modern chaos theory. They are, however, not directly applicable to the homoclinic tangles of equation (2.2). We would therefore move to Hénon maps.

Henon maps are a two parameter family of 2D maps introduced by French astronomer Hénon in 1976 as a simple extension of the quadratic family \( f(x) = 1 - ax^2 \) to 2D. Hénon used a computer to numerically investigate the possible destinations of individual orbits in a parameter range where homoclinic tangles exist. For some parameters he plotted stable periodic orbits, but for others he plotted messy pictures as shown in Figure 4. Smale’s horseshoes, unfortunately, could explain none of the plots.

![Figure 4. Strange attractors in Hénon maps](image)

It took almost two decades for a rigorous dynamics theory to be developed for Figure 4. This theory is mathematically sophisticated and it combines (a) a tour de force analysis from Benedicks and Carleson ([4]); (b) a measure theoretic component rooted in statistical mechanics, initiated first by Sinai ([31]), Ruelle ([30]) and Bowen ([8]) then later developed by Pesin, Ledrappier, Young and others; and (c) further contributions from the work of Benedicks, Young and Viana ([6], [5]). This theory asserts that for the Hénon family there is a positive Lesbegue measure set of parameters, of which the corresponding maps admit no stable periodic orbit. For these maps, individual orbits appear to jump all over the place randomly at any given moment of time. The theory asserts, however, that there is a law of statistics in this seemingly hopeless situation. This is to say that, in an open neighborhood that contains some homoclinic tangles, the asymptotic distribution of points for almost all orbits is the same. A complete dynamical profile for these maps was also provided later by Young and myself ([39]).

Not long after the theory of Benedicks and Carleson on Hénon maps came out, Mora and Viana ([21]) observed that, through a previous theory of Newhouse ([24]), it can be applied to a more general dynamical scenario. Newhouse considered a class...
of homoclinic tangles, in which he replaced the transversal intersection of the two curves in Figure 1(a) by a quadratic tangency. See Figure 5. The maps he studied were a one parameter family and he assumed that the fixed point is dissipative, this is to say that the determinant of the Jacobian matrix at the fix point is with a magnitude < 1. He also assumed that as the parameter varies, the two curves at the point of tangency pass each other.

![Figure 5. Transversal Homoclinic Tangency](image)

In this case, Newhouse observed, there is a small rectangular box close to the point of tangency that would eventually be mapped back into itself, and if we re-scale this box to size \( \approx O(1) \), then the return maps would become virtually a Hénon family. Consequently, there are stable periodic orbits and there is a set of parameters of positive Lesbegue measure for which the Benedicks-Carleson theory on Hénon maps applies. Newhouse also proved that points of transversal homoclinic tangency are structured typically as the intersection of two Cantor sets. Homoclinic tangency is persistent in the sense that the two Cantor sets can not be moved completely apart by a small change of parameter.

As one might have noted, our discussion has moved completely from equation to map. This is a reflection of a historic trend. Theory on maps has become the dominating theme and, for many working on the subject of dynamical systems, equations have faded gradually into the background.

5. The Structure of Homoclinic Tangles

We now bring the study of the homoclinic tangle of equation (2.2) from the back to the foreground. We present a comprehensive description on the overall dynamical structure of the homoclinic tangles for a given equation (2.2) with a two dimensional phase space, recently attained by Ali Oksasoglu and myself ([38]). In this description, results discussed in the previous two sections would fall into their own place as part of a larger and comprehensive picture.

Assume that equation (2.2) for \( \varepsilon = 0 \) admits a homoclinic loop, an open neighborhood of which in the space of \( x \in \mathbb{R}^2 \) we denote as \( U \). The homoclinic tangle of equation (2.2) for \( \varepsilon \neq 0 \), which we denote as \( \Lambda_\varepsilon \), is the collection of all solutions of equation (2.2) that stay in \( U \) for all time. Our purpose is to comprehensively understand the dynamics of \( \Lambda_\varepsilon \) for all small \( \varepsilon \).
One standard step in studying equation (2.2) is to introduce an angular variable $\theta$ to rewrite equation (2.2) as

\[
\frac{dx}{dt} = f(x) + \varepsilon g(x, \theta), \quad \frac{d\theta}{dt} = 1
\]

where $(x, \theta) \in \mathbb{R}^2 \times S^1$ is the extended phase space. Following Poincaré, we study equation (5.1) by iterating maps defined by the solutions. We, however, turn our angle of view away from the direction of the time-$T$ map, studying the homoclinic tangle through a return map in the extended phase space proposed previously by Afraimovich and Shilnikov ([2]). As shown in Figure 6, we start with a short segment intersecting the homoclinic solution in the space of $x$, which we denote as $I$, and we extend $I$ in the direction of $\theta$ to form an annulus $\Sigma = I \times S$ in the extended phase space. The return map we use is defined by the solutions of (5.1) from $\Sigma$ back to $\Sigma$.

Ali and I computed the return map $\mathcal{R} : \Sigma \rightarrow \Sigma$, and attained formulas for $\mathcal{R}$ from equation (5.1) with such details not so attained in the previous literature. The information we extracted from the attained formula is revealing. It has turned out that the seemingly incomprehensible structure of homoclinic tangles as shown in Figure 1(b) is partly intrinsic and partly an artifact of the time-$T$ map.

Our formula for $\mathcal{R}$ revealed that the structure of $\Lambda_\varepsilon$ depends largely on the property of a periodic function, which we could compute explicitly by using the functions on the right hand of equation (5.1) and the given homoclinic solution. This function is an equivalent of the classical Melnikov function. To apply our results to a given equation we need to first verify that all zeros of the Melnikov function are non-tangential. For the reader who knew the Melnikov method for Smale horseshoe ([20]), it should be apparent that we are now imposing a much stronger condition. This is because our objective is the overall structure of the homoclinic tangle $\Lambda_\varepsilon$.

The comprehensive description we extracted from the attained formula for $\mathcal{R}$ is as follows. Starting from $\Sigma$, some solutions would come back to hit $\Sigma$ but others would leave $U$; and we learnt from the formula attained for $\mathcal{R}$ that there are curves roughly vertical in $\Sigma$ that cut $\Sigma$ into a finite collection of vertical strips. Starting from each of these vertical strips the solutions would either all return to $\Sigma$ or all move out of $U$. 
In another word, on each of these vertical strips, the return map $R$ is either defined for all or for none.

The next thing we learnt from the formula for $R$ is that, for each of the vertical strips $V$ where $R$ is defined, the geometry of $R(V)$ is as follows. The strip $V$ is first compressed in the vertical direction and stretched in the horizontal direction, and it is stretched to an infinite length towards both ends. This infinitely long strip is then folded and put back into $\Sigma$. This is to say that, the action of $R$ on $V$ is roughly the same as that of Smale’s horseshoe map, except that the stretched strip is infinitely long. Figure 7 is an illustration for the case in which $\Sigma$ is divided into two strips and the images of $V$ has only one fold. Since $\Sigma$ is an annulus, the infinitely long tails of the image of $V$ would forever wrap around $\Sigma$. In addition, the way $R(V)$ is folded is determined by the shape of the 1D function $\theta + \beta^{-1} \ln M(\theta)$ where $M(\theta)$ is the Melnikov function and $\beta$ is the unstable eigenvalue of the equilibrium solution of equation (2.2).

![Figure 7. The overall structure of homoclinic tangles](image)

Immediately coming out of the description above is a horseshoe of infinitely many branches in $\Lambda_\varepsilon$ for all $\varepsilon$. It takes Smale’s horseshoe of two branches as a subset. A conceptual description of the corresponding invariant set is to forever replace, starting with $V$, every square by a collection of smaller squares inside in an infinite by infinite formation.

The attained formula for $R$ also revealed that, as $\varepsilon \to 0$, $R(V)$ moves in the horizontal direction towards $\theta = +\infty$ with a roughly constant speed $\beta^{-1}$ with respect to $a = \ln \varepsilon^{-1}$. Let us for illustrative purposes talk only of the case of Figure 7. It follows that

(a) There is an intrinsic periodic pattern with a known period in $a = \ln \varepsilon^{-1}$, after which the dynamical behavior of $\Lambda_\varepsilon$ would try to repeat infinitely many times, in a gradually more accurate fashion, as $\varepsilon \to 0$. 
(b) When the folded part of $\mathcal{R}(V)$ is moved out of $V$, which is bound to happen infinitely many times as $\varepsilon \to 0$, the homoclinic tangle $\Lambda_\varepsilon$ contains nothing else but one horseshoe of infinitely many branches.

c) When the folded tip of $\mathcal{R}$ moves to meet its pre-image, which is also bound to happen infinitely many times as $\varepsilon \to 0$, stable periodic solutions of global attractive strength are created in the corresponding $\Lambda_\varepsilon$.

d) The horseshoe of infinitely many branches in all $\Lambda_\varepsilon$ contains infinitely many periodic saddles with roughly vertical stable curves and roughly horizontal unstable curves. The images of the unstable curves would fold, and move through the stable curves, producing numerous values of $\varepsilon$ that admits Newhouse tangency. All results discussed in Section 4 now come into play, where we observe an infinitely complex possibility for Newhouse sinks and Hénon-like attractors.

We caution that all results discussed in Sections 4 and 5 are for equations with dissipation. They do not apply to the energy conservative gravitational systems studied by Henri Poincaré. For conservative systems we can also derive a formula for $\mathcal{R}$, but the results are such that no existing theory on maps are available to offer anything that is nearly as complete as the ones presented in the above for equations with dissipation. It therefore remains a long way for us to even be able to dream about solving the N-body problem.

I would like to thank my friend Kening Lu, who has encouraged and often helped me greatly in the last few years, including in the writing of this essay.

References


[27] H. Poincaré, *Mémoire sur les courbes définies par une équation différentielle*, Journal de Mathématiques, **7** (1881), 375-422, and **8** (1882),251-296.


