A computational procedure to detect a new type of high-dimensional chaotic saddle and its application to the 3D Hill's problem
Waalkens, H.; Burbanks, A.; Wiggins, S.

Published in:
Journal of Physics A: Mathematical and Theoretical

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2004

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
LETTER TO THE EDITOR

A computational procedure to detect a new type of high-dimensional chaotic saddle and its application to the 3D Hill’s problem

H Waalkens, A Burbanks and S Wiggins

School of Mathematics, Bristol University, University Walk, Bristol, BS8 1TW, UK

E-mail: H.Waalkens@bris.ac.uk

Received 9 February 2004, in final form 29 April 2004
Published 2 June 2004
Online at stacks.iop.org/JPhysA/37/L257
doi:10.1088/0305-4470/37/24/L04

Abstract

A computational procedure that allows the detection of a new type of high-dimensional chaotic saddle in Hamiltonian systems with three degrees of freedom is presented. The chaotic saddle is associated with a so-called normally hyperbolic invariant manifold (NHIM). The procedure allows us to compute appropriate homoclinic orbits to the NHIM from which we can infer the existence of a chaotic saddle. It also allows us to detect heteroclinic connections between different NHIMs. NHIMs control the phase space transport across an equilibrium point of saddle-centre-centre-centre stability type, which is a fundamental mechanism for chemical reactions, capture and escape, scattering, and, more generally, ‘transformation’ in many different areas of physics. Consequently, the presented methods and results are of broad interest. The procedure is illustrated for the spatial Hill’s problem which is a well-known model in celestial mechanics and which gained much interest, e.g. in the study of the formation of binaries in the Kuiper belt.

PACS numbers: 05.45.Jn, 45.50.Pk

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Chaotic saddles are the saddle-type invariant Cantor sets associated with a horseshoe construction [1]. They play a central role in many complex dynamical phenomena, e.g., the existence of supertransients [2] and the fractal structure of chaotic scattering [3]. The chaotic saddles constructed to date have been related either to homoclinic orbits associated with hyperbolic periodic orbits or to special types of equilibria. In this letter, we are concerned
with a fundamentally new type of high-dimensional chaotic saddle structure that can occur only in Hamiltonian systems with three or more degrees of freedom (DOFs), and we describe a computational method for detecting it which also illustrates its geometrical complexity.

The chaotic saddle is associated with a normally hyperbolic invariant manifold (NHIM) [4]. The physical significance of NHIMs arises from the fact that their stable and unstable manifolds control the phase space transport across equilibria of saddle-centre-centre stability type. This is not only the fundamental mechanism for the evolution from reactants to products in chemical reactions, but also for ‘transformations’ in general in a large, and diverse, number of applications including, e.g. ionization problems in atomic physics [5], rearrangements of clusters [6], cosmology [7] and solid state and semi-conductor physics [8, 9]. Though it had been recognized that it is important to understand the dynamics near saddle-centre-centre equilibria it was only recently that new developments in dynamical systems theory offered the theoretical framework and computing power offered the means to study the phase space structure near saddle-centre-centre centres for systems with three or more DOFs [10–13]. Besides the NHIM and its stable and unstable manifolds it is now possible to compute a codimension 1 submanifold of the energy surface, the so-called transition state, which is transverse to the Hamiltonian flow and locally divides the energy surface into two disjoint components. Locally, the transition state provides the only means of passing from one phase space region (associated with ‘reactants’) to another phase space region (associated with ‘products’), i.e. trajectories must cross the transition state in order to ‘react’. The transversality of the transition state to the Hamiltonian flow is essential for rate calculations as it eliminates the problem of locally recrossing trajectories.

In this letter, we go beyond this local result and consider more global issues associated with the dynamics related to the transition state and NHIM. In particular, we describe a computational method for determining the existence of homoclinic and heteroclinic connections to NHIMs. Using recent results in [14], the existence of certain types of these homoclinic orbits allows us to infer the existence of a new type of chaotic saddle (in fact, a Cantor set of chaotic saddles). Because of the ubiquity of saddle-centre-centre-type equilibria in applications (as described above), we expect the methods and results presented here, which are applicable to three and more DOFs, to be of broad interest.

The physical system that we choose to illustrate our method is one from celestial mechanics; the 3D Hill’s equations [15]. Advances in detector technology have opened up new frontiers in celestial mechanics with the discovery of trans-Neptunian objects and binary systems in the Kuiper belt. The calculation of capture probabilities requires the study of the 3D Hill’s equations (rather than the thoroughly studied 2D case) because many of the capture events occur from high inclination (see e.g., the recent work by Goldreich et al [16]).

2. Hill’s problem and the phase space structure near saddle-centre-centre equilibria

The circular restricted three-body problem (CRTBP) models the motion of a tiny particle under the gravitational influence of one (large) primary mass and one (smaller) secondary mass both in circular orbits about their common centre of mass [17]. Hill’s problem is a limit version of the CRTBP which describes the motion of the particle in a neighbourhood of the secondary mass. Hill’s equations can be derived from the following Hamiltonian in dimensionless units,

\[ H = \frac{1}{2} (p_x^2 + p_y^2 + p_z^2) + yp_x - xp_y - x^2 + \frac{1}{2} (y^2 + z^2) - \frac{3}{r} \]

where \( r = (x^2 + y^2 + z^2)^{1/2} \). It is well known that Hill’s equations have two equilibria, which are denoted traditionally by \( L_1 \) and \( L_2 \). The matrices associated with linearizing Hamilton’s
equations about each equilibrium have a pair of real eigenvalues of equal magnitude and opposite sign \((\pm \sqrt{2\sqrt{7} + 1})\) and two pairs of pure imaginary complex conjugate eigenvalues \((\pm i2, \pm i\sqrt{2\sqrt{7} - 1})\). This means that \(L_1\) and \(L_2\) are equilibria of saddle-centre-centre type. A detailed theory for phase space transport near saddle-centre-centre equilibrium points has been developed in recent years [10–13]. For energies slightly above that of the saddle-centre-centre equilibrium point, on each five-dimensional energy surface there exists an invariant three-dimensional sphere \(S^3\) of saddle stability type. This 3-sphere is significant for two reasons:

- It is the ‘equator’ of a four-dimensional sphere called, in much of the literature, the transition state. Except for the equator (which is an invariant manifold), the transition state is locally a ‘surface of no return’ in the sense that trajectories which have crossed the transition state must leave a certain neighbourhood of the transition state before they can possibly cross the transition state again. For energies ‘sufficiently close’ to the energy of the saddle-centre-centre equilibrium point, the transition state satisfies the bottleneck property. This means that the energy surface locally has the geometrical structure of \(S^4 \times I\) (i.e., 4-sphere \(\times\) interval) and the transition state divides the energy surface into two disjoint components. Moreover, the only way a trajectory can pass from one component of the energy surface to the other is to pass through the transition state.

- The 3-sphere is a normally hyperbolic invariant manifold [4], which means that the expansion and contraction rates of the dynamics on the 3-sphere are dominated by those transverse to it. Just like a ‘saddle point’ it therefore has stable and unstable manifolds. In this case the stable and unstable manifolds are four dimensional, having the structure of spherical cylinders, \(S^3 \times \mathbb{R}\). Hence, they are of one less dimension than the energy surface and act as ‘ separatrices’; they ‘enclose’ a volume of the energy surface. Their key dynamical significance is that the only way that trajectories can pass through the transition state is if they are inside the region of the energy surface enclosed by the stable and unstable spherical cylinders.

These phase space structures can be realized through a procedure based on Poincaré–Birkhoff normalization by which explicit formulae for the NHIM, its stable and unstable manifolds and the transition state are given [13] in ‘normal form coordinates’. The phase space structures are then mapped back into the original coordinates by the inverse of the normal form (NF) transformation.

It is important to understand how the normal form is used to compute phase space structures since, in general, we do not expect the normal form expansion to converge (see [18–22], and [23] for an overview). The goal is to obtain a neighbourhood of the saddle-centre-centre equilibrium point in phase space that is as large as possible, in which the approximation resulting from the truncation of the NF (to some finite order) yields the ‘desired accuracy’. What we mean by ‘desired accuracy’, and how we determine it, will be explained in a moment. This ‘neighbourhood of validity’ of the NF in phase space has to be large enough to contain the transition state (and hence the NHIM which is itself contained in the transition state) for the energy of interest. The larger the neighbourhood of validity, the higher one can go in energy above the energy of the saddle-centre-centre in order to compute the transition state, the NHIM, and the local parts of its stable and unstable manifolds. For a given neighbourhood of an equilibrium point, we take the NF computation to successively higher orders until either (i) no improvement in accuracy is seen with increasing order, (ii) the desired accuracy is reached or (iii) we reach practical limits on the size of our computations. For fixed neighbourhoods about Hill’s equilibria \(L_1\) and \(L_2\) and up to the maximum order that we can currently reach.
with our code for the NF computation, we observe that the accuracy of the NF still improves as the degree of the NF is increased. Using the transformations back and forth between the NF coordinates and the original phase space coordinates, the accuracy of the NF is determined by a battery of checks which include the conservation of the original Hamiltonian on the computed transition state (which contains the NHIM), invariance of the computed NHIM under the original Hamiltonian vector field, and conservation of the integrals resulting from the NF along trajectories computed by integrating the original equations of motion as they pass through the neighbourhood under consideration. The NFs that we have used here satisfy all these checks to a high accuracy, as we will note later.

3. Computation and visualization of the phase space structures near $L_1$ and $L_2$

We will visualize the phase space structures near Hill’s equilibria $L_1$ and $L_2$ as their projections to configuration space. In understanding the result, and its implications, it is useful to recall that the level sets of the effective potential energy

$$V \equiv H - \frac{1}{2} |(p_x + y, p_y - x, p_z)|^2 = -\frac{3}{2} x^2 + \frac{1}{2} z^2 - \frac{3}{r}$$

the so-called zero velocity surfaces, confine the motion in configuration space.

In figure 1 we show the zero velocity surface (ZVS) $V = E$ for an energy $E = -4.4$ which is 0.1 above the energy of the equilibrium points. The ZVS encloses the energetically allowed volume of configuration space. For Hill’s equations its shape has two ‘bottlenecks’ associated with $L_1$ and $L_2$ which divide the allowed volume into three regions: a bounded region about the origin (for which $-1 < x < +1$ (A)) and two unbounded regions (for which $x < -1$ (B) and $x > +1$ (C), respectively). The phase space structures near $L_1$ and $L_2$ regulate the transport through the bottlenecks between these regions.

The NFs about $L_1$ and $L_2$ are related by symmetry. It is thus sufficient to compute explicitly only the NF about $L_1$. As a result of the two complex eigenvalues associated with the equilibria
being rationally independent, the NF, to any desired finite order of computation, is completely integrable with integrals given by
\[ I = p_1q_1, \quad J_i = \frac{1}{2}(q_i^2 + p_i^2), \quad i = 2, 3, \]
where the \((q_i, p_i)\) are the (canonically conjugate) pairs of coordinates for the normal form. The corresponding Hamiltonian can be written solely as a function of the integrals,
\[ H = H(I, J_2, J_3), \]
and Hamilton’s equations decouple into the product of independent linear systems
\[ \dot{q}_1, \dot{p}_1 = \frac{\partial H}{\partial I}(q_1, -p_1) \quad (q_i, p_i) = \frac{\partial H}{\partial J_i}(p_i, -q_i) \quad i = 2, 3. \]

The transition state 4-sphere is given by \( q_1 = p_1 \). From the equations of motion it is easy to see that it is a ‘surface of no return’. Its equator \( p_1 = q_1 = 0 \) is the NHIM (3-sphere). It has four-dimensional stable \((q_1 = 0)\) and unstable \((p_1 = 0)\) manifolds. In the normal form coordinates, in the energy surface volume enclosed by the stable and unstable manifolds, \( \mathcal{I} \) is positive; outside, \( \mathcal{I} \) is negative. The NHIM has a special structure; it is foliated by a one-parameter family of invariant 2-tori (the ‘Hopf fibration’) \([10, 11, 13]\). These tori can be parametrized by e.g. the integral \( J_2 \) where \( J_3 \) is then given implicitly by energy conservation and \( J_2 = 0 \). At the minimal and maximal values of \( J_2 \) the 2-tori degenerate to periodic orbits, the so-called ‘Lyapunov orbits’. The 2-tori and the periodic orbits have three-dimensional and two-dimensional stable and unstable manifolds, respectively, which are contained in the four-dimensional stable and unstable manifolds of the NHIM. These geometrical considerations will play an important role as described below.

We perform the NF computation, using the computer-algebra system Mathematica, up to degree 18 (i.e. the NF Hamiltonian may be written as a sum of homogeneous polynomials of degree up to 18 in the normal form coordinates). As a result, the NF Hamiltonian is a sum over 219 multivariate monomials in \((I, J_2, J_3)\). Each component of the mapping between NF coordinates and the original phase space coordinates involves sums over about 50,000 multivariate monomials. The resulting transition state and NHIM, for the chosen energy \( E = -4.4 \), are shown in the original coordinates in figure 1. Note that the transition state blocks completely the ‘bottleneck’ in the ZVS in configuration space; it also blocks the bottleneck in the corresponding level set of the Hamiltonian in phase space.

**4. Computation of connections homoclinic to, and heteroclinic between, the NHIMs near \( L_1 \) and \( L_2 \)**

The ‘saddle integral’, \( \mathcal{I} = p_1q_1 \), plays the key role in our numerical approach to detecting orbits connecting the same NHIM, or orbits connecting different NHIMs. The procedure is as follows. Taking into account the Hopf fibration of the NHIM, choose in the NF coordinates an invariant 2-torus on the NHIM. Displace this 2-torus slightly in the direction of its unstable manifold \((p_1 = 0, q_1 = \epsilon)\).

- Choose initial conditions on the shifted 2-torus and map the initial conditions back into the original coordinates using the NF transformation.
- Integrate the initial conditions forward in time using Hill’s equations. Since they are in the unstable manifold they will leave the neighbourhood in which the NF transformation is valid (which is why we integrate them in the original coordinates).
- Check if the trajectory returns to the neighbourhood of \( L_1 \) or \( L_2 \) where the NF is valid. If so, map it into the NF coordinates and check the value of its saddle integral \( \mathcal{I} \).

If the saddle integral \( \mathcal{I} \) were zero, then the trajectory must be on the stable manifold of the respective NHIM (since the trajectory is already on the unstable manifold and two unstable manifolds cannot intersect). The procedure to find homoclinic and heteroclinic
Figure 2. The top panels highlight individual 2-tori in the Hopf fibration of the NHIM near $L_1$ that (when displaced along the unstable manifold) generate the initial conditions for our computational method. Also shown are contours of values of the saddle integral $I_{L_1}$ and $I_{L_2}$, obtained after the initial conditions are integrated into a neighbourhood of $L_1$ or $L_2$, respectively, on these 2-tori. The 2-tori are for $J_2 = n J_{2\text{max}}/4$, $n = 1, 2, 3$, where $J_{2\text{max}}$ is the maximum $J_2$ on the NHIM. $n = 0$ and $n = 4$ correspond to the two Lyapunov periodic orbits. The 2-tori are parametrized by the angles $\alpha_2$ and $\alpha_3$ conjugate to $J_2$ and $J_3$. For clarity the bottom panels show the tori in the covering space. ($E = -4.4$.)

connections consists of varying the initial conditions on the shifted 2-torus to find the zeros of $I$. The procedure can be understood as a shooting method between the $I$-fibres of the locally valid NFs. We apply the method to Hill’s problem by displacing ($\epsilon = 10^{-4}$) initial conditions on invariant 2-tori of the NHIM near $L_1$ along the respective unstable manifold branches which are directed towards the bounded region about the origin, see figure 1. These initial conditions are propagated by integrating Hill’s equations where the singularity at the origin is taken care of by the Kustaanheimo–Stiefel regularization [24]. If the resulting trajectory reenters or enters the neighbourhood of validity of the NF about $L_1$ or $L_2$, respectively, we check the value of the integral $I$. A positive $I$ means that the trajectory will cross the respective transition state which leads to an exit to the outside of the bounded region about the origin. In this case the integration is stopped. A negative value of $I$ means that the trajectory does not exit on this approach of $L_1$ or $L_2$ and the integration is continued until the trajectory reaches a validity neighbourhood of the NF with positive $I$. It is to be noted that along the part of a trajectory which traverses the validity neighbourhood of the NF the values of the integrals $(I, J_2, J_3)$ are conserved to 12 digits (and even more when closer to the NHIM), demonstrating the high accuracy of the NF.

We illustrate the results in figure 2 where contours of the integral $I_{L_1}$ (light blue/dark blue; corresponding to exiting through the transition state near $L_1$) and $I_{L_2}$ (yellow/red; corresponding to exiting through the transition state near $L_2$) are shown plotted on the corresponding torus of initial conditions. At the boundary of a light blue/dark blue region $I_{L_1}$ goes to zero and points on the boundary indicate the existence of homoclinic orbits which connect the NHIM near $L_1$ to itself. Similarly, the boundary of each yellow/red region indicates heteroclinic connections between the NHIMs near $L_1$ and $L_2$.

Though the individual regions in figure 2 are themselves regular, in the sense that they each have a smooth boundary, the disposition of the regions is very intricate. We illustrate this by
repeating the procedure described above for initial conditions on the two-dimensional unstable manifold of the Lyapunov periodic orbit near $L_1$ which has $J_3 = 0$ and in configuration space lies in the $(x, y)$-plane. Integrating these initial conditions forward in time, if they return to a neighbourhood of $L_1$ or $L_2$ where the NF is valid we plot the value of $I_{L_1}$ in blue or $I_{L_2}$ in red as a function of the angle conjugate to $J_2$ along the Lyapunov periodic orbit. These are shown in the top panel of figure 3. Since these functions are highly oscillatory, for visualization purposes we also colour the region under the value of the function with the corresponding colour. Each zero of $I_{L_1}$ or $I_{L_2}$ indicates the existence of an initial condition which in time is backward asymptotic to the Lyapunov periodic orbit and forward asymptotic to the NHIM near $L_1$ (blue) or the NHIM near $L_2$ (red), respectively. The two panels below the top panel in figure 3 represent successive magnifications of smaller regions of figure 3. From these one sees a self-similar structure well known from classical scattering systems.

5. Homoclinic connections and chaotic saddles

While orbits homoclinic to normally hyperbolic invariant tori have been studied ([25]), tori cannot be normally hyperbolic in Hamiltonian systems [26]. The NHIM is normally hyperbolic, yet there are no theorems describing the dynamics associated with orbits homoclinic to a normally hyperbolic invariant sphere (part of the difficulty here comes from the fact that a sphere cannot be described by a single coordinate chart, but see [7] for numerical evidence that this could be an important mechanism for chaos). An important piece of this problem has recently been solved by Cresson [14] who proved that if the stable and unstable manifolds of a torus in the NHIM intersected transversely, then there exists a (uniformly) hyperbolic invariant Cantor set on which the dynamics is conjugate to a shift map, i.e., a chaotic saddle. The stable and unstable manifolds of a 2-torus are three dimensional in the five-dimensional energy surface. A transverse intersection is necessarily one dimensional, i.e., a trajectory.
When the effect of the neglected terms in the normal form expansion (i.e., the non-normalized ‘tail’ of the expansion) are included we expect, by KAM theory, that a Cantor set of non-resonant tori in the Hopf fibration will persist. Homoclinic intersections of the stable and unstable manifolds of a given 2-torus (rather than more general connections between the manifolds of such a torus and those of the entire NHIM) can be detected easily by our numerical procedure described above. We therefore not only check the value of the integral $I$ upon entering the neighbourhood of validity of the same NF that we started from, but also the difference between the centre integral $J_2$ of the 2-torus we started from and the centre integral $J_2'$ in the entered validity neighbourhood. If both $J_2 - J_2'$ and $I$ vanish we have a homoclinic orbit to the 2-torus under consideration. The result of such a procedure is shown in figure 4(a) where homoclinic connections appear as the intersection points of the zero contours of $J_2 - J_2'$ (sharp boundaries between dark blue and light blue within the blue regions) and those of $I$ (boundaries of the blue regions themselves). Figure 4(b) shows an example of a homoclinic orbit which corresponds to one of the prominent intersection points marked by a black dot in the figure. In fact there are many more intersection points located in the finer structures of the left part of figure 4(a). There is a tendency that the homoclinic orbits become more complicated in the finer regions.

6. Conclusions

In this letter, we have described a numerical method for constructing homoclinic orbits to, and heteroclinic orbits between, NHIMs in Hamiltonian systems with three DOFs. Our method thereby detects a Cantor set of chaotic saddles by constructing appropriate homoclinic orbits (namely, those for single tori in a NHIM). We have illustrated this for the spatial Hill’s problem. The implications for this problem should be of current interest in celestial mechanics. Recently, it has been shown that chaos plays an important role in the energetics of capture [27], and this Cantor set of chaotic saddles should be central to this process. Chaotic scattering in systems with three or more DOFs is poorly understood, and our results and methods should provide a window into this subject.

Concerning transition state theory, as described by Truhlar [29], there are two types of trajectories that exhibit multiple recrossings of the transition state: local and global. As
described in the introduction, the correct choice of transition state solves the local recrossing problem. However, characterizing the global recrossing problem, i.e. the question of whether there are trajectories, which, after crossing the transition state and leaving its neighbourhood, return to the transition state and cross it again, is more difficult. Under additional assumptions on the geometry of the homoclinic or heteroclinic orbits, the chaotic saddles consist of trajectories that enter and leave a neighbourhood of the transition state, recrossing it infinitely often. This will be the subject of a future publication.

It would also be of some interest to actually compute and visualize the chaotic saddles themselves. A recently developed numerical method [28] may play a role for this purpose. Finally, we remark that our results are not limited to three DOFs. The general theory developed in [10–13] applies to Hamiltonian systems with arbitrary DOFs. The numerical method that we present here for finding homoclinic and heteroclinic connections generalizes easily since the saddle integral is a scalar function regardless of the number of DOFs.

Acknowledgments

We thank J Palacián and P Yanguas for Mathematica code which we modified for the NF computation. This work was supported by the Office of Naval Research, grant no N00014-01-1-0769 and the Royal Society. HW acknowledges support from the Deutsche Forschungsgemeinschaft, grant no Wa 1590/1-1.

References

[3] See the Chaos Focus issue Chaotic Scattering in Chaos 34