STATISTICAL STABILITY OF THREE AND MORE BODY HIERARCHICAL SYSTEMS IN CELESTIAL MECHANICS

by

ALASTAIR JAMES CALUM McDONALD B.Sc.

Thesis

submitted to the

University of Glasgow

for the degree

of Ph.D.

Department of Astronomy,
The University,
Glasgow G12 8QQ

June 1986

To my mother and father,

in gratitude for 25 years

of love and support

C O N T E N T S

PREFACE

Then to the Heav'n itself I cried,
Asking, "What Lamp had Destiny to guide
Her little Children stumbling in the Dark?"
And - "A blind Understanding!" Heav'n replied.

Rubáiyát of Omar Khayyám

It seems that not everybody in Persia in the eleventh century was as convinced as the astrologers that the movements of the heavens controlled the destiny of Man. Nevertheless, for many centuries before and since, kings and emperors rewarded handsomely those astronomer/astrologers who could give them advice based on the movements of the planets and other celestial bodies. (There may be some astronomers today who would wish for similar generous patronage). Since the advent of modern celestial mechanics with the work of Isaac Newton, orbital motion has been studied for its own sake and, in the last thirty years, for the purposes of sending artificial satellites and manned craft into space. Yet for 300 years, one of the most important questions posed by celestial mechanics remains unanswered: are the motions of the planets in the Solar System stable? Could planets collide or even escape? Countless workers since Newton's time have sought Lamps to the Destiny of the Solar System, but our Understanding is still obscured by many blind-spots.

This thesis does not claim to give any definitive answers to these questions. It does indicate how to obtain quantitative estimates of the likelihood of certain events occurring. Simple statistical

methods are applied to the results of numerical experiments and give probabilities of planetary orbits crossing or bodies escaping dynamical systems altogether.

In Chapter 1 a general review of the problem of the Solar System's stability is given along with brief descriptions of methods and definitions of stability which have been used in the past. This thesis studies the stability of real and fictitious dynamical systems not necessarily associated with the Solar System. It investigates one particular definition of stability, namely *hierarchical stability*, using special perturbation methods. The definitions of hierarchical systems, hierarchical stability and empirical stability parameters are reviewed in Chapter 2. These will form the basis for subsequent numerical experiments.

One further definition of stability - *Hill stability* is an important condition for hierarchical stability. It has been studied in a mathematically rigorous way in the problem of three massive bodies in mutually perturbed orbits. This analysis as well as some new numerical results are given in Chapter 3.

Numerical integration experiments were carried out, with the aid of a mainframe computer, to study the period of time for which various three-body systems remain stable. Several hundred fictitious systems with different masses and starting conditions were studied. In each case, all three bodies' orbits lay in the same plane. In some systems, all the bodies orbited in the same direction (direct); for other systems, one body orbited in the opposite direction from the other two (retrograde). The results of these experiments are presented in Chapter 4 (for retrograde systems) and Chapter 5 (for direct systems). The results are grouped in such a way that analytical curves may be

fitted to the data. This allows predictions of stability lifetimes for similar systems without the need for lengthy numerical integration experiments. Systems whose masses, initial positions and initial velocities fall into certain ranges are always stable. These regions of hierarchical stability are mapped out and compared with corresponding regions of Hill stability. In the case of direct systems, commensurabilities give rise to large fluctuations in stability lifetimes, if the initial conditions are varied slightly. Additional statistical methods are described in Chapter 5 to cope with this effect.

In Chapter 6, the results of Chapters 4 and 5 are compared with real three-body systems within the Solar System. Possible origins of the Solar System are discussed in the light of the results.

In Chapter 7, four and more body systems are examined for alignments of the bodies in their orbits (syzygies). A statistical analysis of the numbers of syzygies occurring in a given time leads to the discovery of resonances in the orbital frequencies. The theory is developed and applied to the results of numerical experiments.

Chapter 8 briefly states some of the questions that have not been considered in this work as well as some new ones that have arisen from it.

Appendix A is a set of useful statistical tables. Appendix B is a discussion of some limiting cases of Hill stability in the general three-body problem. Appendix C gives a detailed mathematical background to the statistical methods used in Chapter 5.

The original work of this thesis is contained in Chapters 4 to 7, the second half of Chapter 3, and Appendices B and C. Results from Chapter 5 have been published in *Stability of the Solar System and Its Minor Natural and Artificial Bodies*, in the *NATO ASI Series*.

Results from Chapter 7 have been accepted for publication in *Celestial Mechanics*. Parts of Chapters 3 to 6 are in preparation as papers.

This work does not answer any questions about the ultimate stability of the Solar System. I shall be happy though if it helps to irradicate even one blind-spot in our understanding of the Solar System's destiny. I leave the study of Man's destiny to others. (Besides, we Scorpios don't believe in astrology).

ACKNOWLEDGEMENTS.

The work in this thesis was carried out while the author was a research student in the Department of Astronomy, University of Glasgow. I extend my thanks to all the staff and students of the department, past and present, who have made my stay so interesting and enjoyable.

Even with his experience of psychic phenomena, it is unlikely that Professor Archie Roy could have foreseen that the schoolboy who visited him in 1977 would later become one of his research students. His enthusiasm for astronomy in general and celestial mechanics in particular proved highly infectious and is probably incurable. It has been a pleasure to work closely with Dr. Ian Walker over the past three years. I have benefitted greatly from his experience of celestial mechanics and computing. The jokes and whistling in two part harmony were enjoyed by both of us, if not by the rest of the department. I thank both Archie and Ian for their time and experience, but most of all their friendship.

Other members of the department deserve special mention. Thanks to Professor P.A. Sweet, for help and encouragement over seven years. I am only sorry I was not able to profit by spending more time with him. Thanks: to Professor John Brown and Dr. John Simmons, for getting me started; to Dr. Alex MacKinnon for discussions on maths, physics, music and "IF"; to Dr. Brian Stewart for discussions on statistics; to Mr. Alan Thompson and Miss Vivette Carlaw, for suffering 18 months of puns and whistling; to Miss Bonnie Stevis, for confusing her with information in order to get my own thoughts straight; to Mrs. Margaret Morris, whose artistic talents are wasted on diagrams. To everyone, thank you for all your kindness.

Due to my supervisor's consideration, I have been able to participate in numerous meetings with other workers outside the department. I have enjoyed fruitful discussions with Dr. Anna Nobili, Dr. Andrea Milani, Dr. Giovanni Valsecchi and Dr. Andrea Carusi. My thanks also go to the members of the LONGSTOP project and the participants at the NATO ASI conference in Cortina 1984, particularly Dr. Christian Marchal. I have benefitted from meeting them all.

I would particularly like to thank Mrs. Lillias Williamson for relieving me of the burden of typing this thesis and for her continuing kindness to one of "her boys".

My thanks go to the staff of the Computing Service at the University for running hundreds of overnight jobs at the 98% confidence level and to the Science and Engineering Research Council for providing me with a grant to carry out this research.

Finally, thanks go to all my friends and family, especially my sister Eileen, for their interest and encouragement. Much of this thesis was written while staying with my aunt and uncle, Irene and Arthur Gurr. Their kindness has played a most important part in its completion and I cannot thank them enough. Lastly, to my parents Bunty and Bill - the dedication says it all.

Alastair J.C. McDonald

STATISTICAL STABILITY OF THREE AND MORE BODY HIERARCHICAL SYSTEMS IN CELESTIAL MECHANICS

SUMMARY

SUMMARY

After 300 years, celestial mechanics is still unable to say whether or not the orbits of the planets and satellites in the Solar System are stable. The studies that have been performed along with the observational evidence strongly suggest that the majority of orbits are stable, including all planetary orbits. However a definitive answer is still not available. One of the main objectives of this thesis is to obtain statistical estimates of the likelihood of particular orbits being stable.

Many dynamical systems in nature can be defined as *hierarchical*; where the bodies' mutual separations can be ordered and that ordering (hierarchy) is preserved for a time considerably longer than any of the periods of revolution. The equations of motions are expressed in Jacobian coordinates. An expansion of the force function yields a set of dimensionless parameters, the *empirical stability parameters*, which represent the perturbations on the osculating Keplerian ellipses.

This thesis investigates the stability of coplanar hierarchical systems. Particular consideration is given to coplanar initially circular, three-body systems, which can be divided into two classes according to their rotational sense: direct and retrograde.

There are many definitions of stability. This study uses the concept of *hierarchical stability*, i.e. no bodies escape the system; there are no secular changes in the semi-major axes, eccentricities and inclinations defining the osculating orbits; the hierarchy of the system is preserved. This last condition is guaranteed if *Hill-type stability* can be proved. An analytical theory exists for the Hill-type stability of three bodies, which makes use of the topology

of the zero velocity curves. In this theory, it is found that direct systems are stable for a wider range of initial conditions than retrograde systems.

Several hundred numerical integration experiments have been performed for a variety of fictitious direct and retrograde three-body systems. The osculating elements of certain retrograde systems are studied in detail to show the differences between stable and unstable behaviour.

For unstable systems, the times until instabilities have appeared are noted. The resulting data, which compares these stability lifetimes against the initial relative separations of the bodies, is grouped into classes of equal empirical stability parameters. A curve is fitted to each class of data. For retrograde systems, the data varies smoothly and the curves are in good agreement. This allows subsequent predictions of stability lifetimes with good accuracy. The existence of commensurabilities produces wide fluctuations in the data for direct systems, so the curves do not fit so well. Additional statistical techniques are employed to derive probable ranges of stability lifetimes without the need for numerical integration.

The results of these numerical experiments indicate that retrograde systems are stable for a wider range of initial conditions than direct systems. This contradicts the Hill-type stability theory. The numerical results are compared with real three-body subsystems of the Solar System. They imply that the observed lack of retrograde systems may be due to the manner of the origin of the Solar System; not to any stability selection effect. Possible origins of the Solar System are discussed.

Four and more body systems are discussed with reference to an

orrery model of hierarchical, coplanar, unperturbed circular orbits. A theory is developed which predicts the average period of occurrence of near *syzygies* (alignments of the bodies). The theory is found to be in agreement with numerical simulations. Commensurable systems are characterised as having a period of syzygy occurrence which is radically different from that predicted. Examples are given of commensurable systems within the Solar System that are detected by this method. This method can be used to search for critical arguments using apsidal as well as bodily syzygies. This would best be achieved using a numerical integration method with the orrery model acting as interpolator. *Mirror configurations* could also be found leading to a discovery of near periodic motion in the Solar System.

CHAPTER 1

STABILITY OF DYNAMICAL SYSTEMS IN CELESTIAL MECHANICS

1.1 The Structure and Stability of the Solar System

The Earth is but one body in the thousands that make up the Solar System. The principal object is the Sun, a star which is a thousand times more massive than Jupiter, the second largest body. Orbiting the Sun are nine major planets (Figure 1.1) many having their own systems of satellites.

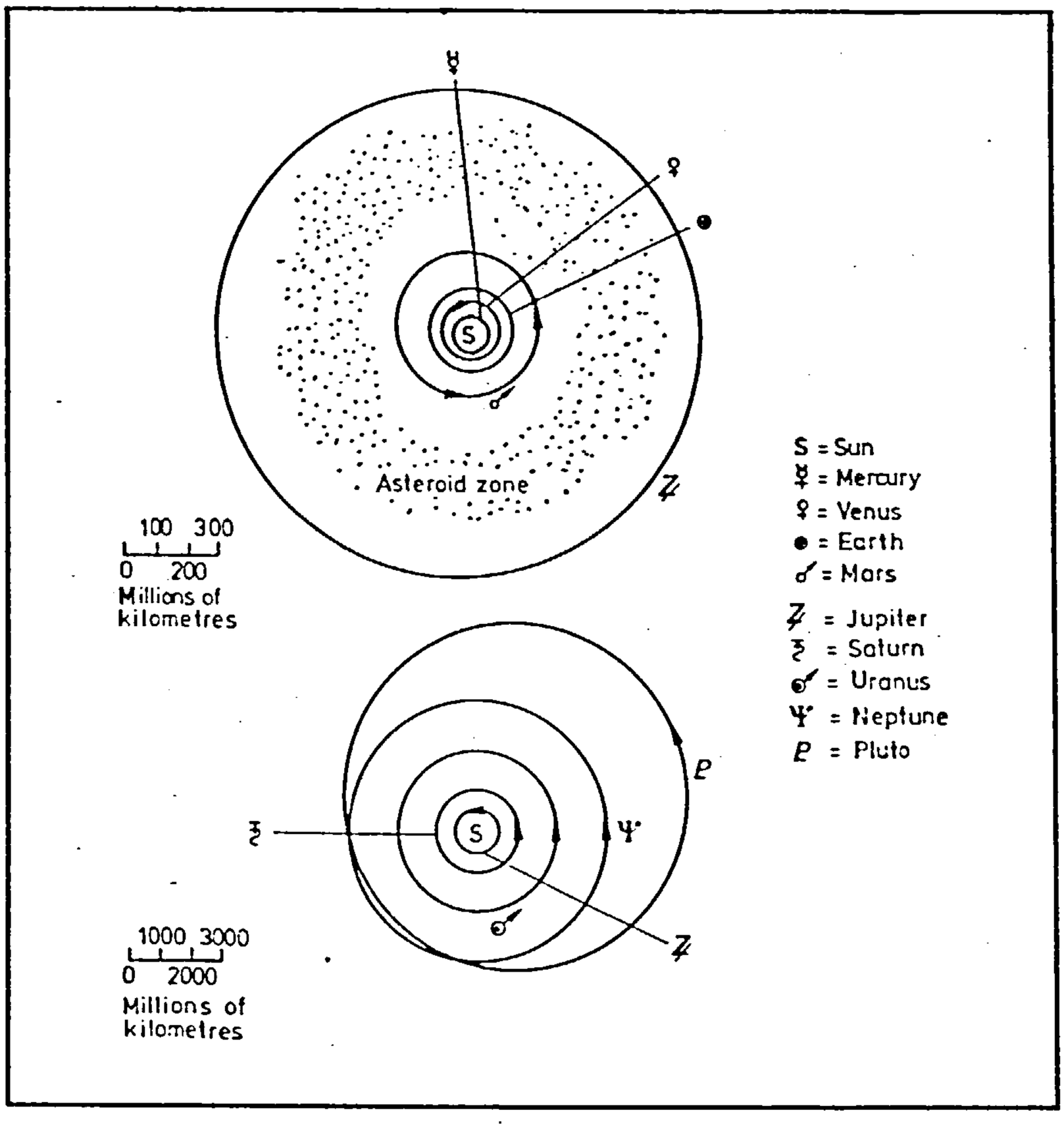


Figure 1.1: Orbits of the major planets in the Solar System

There are over forty major satellites, (although the definitions of major and minor satellites become increasingly hazy as more small bodies are discovered by the Voyager mission and ever improving ground-based observations). Minor bodies include the asteroids, comets and meteor

1.1

streams. On the very small scale we must consider the planetary rings and the interstellar medium.

In terms of composition, the planets may be divided into two groups. Firstly, there are the gas giants; Jupiter, Saturn, Uranus and Neptune. As implied, these planets are much larger than the other five and are mainly comprised of hydrogen and helium. All have their own satellite systems and three of them have been observed to have rings. It may well be that the fourth, Neptune, also possesses rings but these have not yet been detected. The second group is composed of the terrestrial planets; Mercury, Venus, Earth, Mars and Pluto. These planets are much smaller and are mostly composed of silicate material. They have fewer satellites than the gas giants, Mercury and Venus having none whatsoever.

Kepler was able to describe the motions of the planets by his three famous laws:-

(i) The orbit of each planet is an ellipse with the Sun at one focus.

(ii) The rate at which the radius vector from Sun to planet sweeps area is constant.

(iii) The cubes of the semi-major axes of the planetary orbits are proportional to the squares of the planets' orbital periods.

These three laws are exact if we assume that the bodies are points and that the planets do not gravitationally disturb each other. In reality neither assumptions are true but Kepler's laws still give, in most cases, very good approximations to both planetary motion about the Sun and satellite motion about the planets.

Most of the planets' orbits are inclined within a few degrees of

1.1

the Earth's orbit and possess near elliptical orbits with low eccentricities. The two exceptions are Mercury (eccentricity $\sim$ 0.21, inclination $\sim 7^{o}$) and Pluto (eccentricity $\sim$ 0.25, inclination $\sim 17^{o}$). Without exception, all the planets orbit in the same direction about the Sun. Most of the planets rotate on their axes in the same directions as they move in their orbits, the two exceptions being Venus, whose equator is inclined at $\sim 179^{o}$ to its orbit, and Uranus, whose equatorial inclination is $\sim 98^{o}$.

In general the planets' orbits are well spaced and can be unambiguously ordered in increasing size, i.e. in a hierarchical arrangement (Figure 1.1). The exceptions are Neptune and Pluto whose orbits cross. However, Pluto's highly inclined orbit prevents the two bodies from getting too close to each other and there is evidence from numerical experiments to suggest that additional dynamical mechanisms may be present which preserve this state, (see Section 1.6).

The planetary distances from the Sun may be roughly described by Bode's law, namely

$$r_n = 0.4 + 0.3\ (2^n)$$

where r_n is measured in units of the Sun-Earth distance. For Mercury, $n \to -\infty$; Venus , $n = 0$; Earth, $n = 1$; and so on, including the asteroids at $n = 3$. This empirical result works well out to Uranus, but is poor for Neptune and fails for Pluto. Because of its lack of physical justification, many believe Bode's law to be coincidental. However similar laws can be found for the larger satellite systems, so the controversy continues.

The satellite systems, while similar to the planetary system but

1.1

on a smaller scale, show a greater variety of behaviour. Many of the satellite orbits have high eccentricities and inclinations. At least two satellites are spiralling in, towards their planets. These are Phobos aroumd Mars and Triton around Neptune. In both cases this is due to tidal deformation of the planet acting as a brake on the satellite, (Section 6.5). Some satellites are observed to be orbiting in the opposite direction from the other bodies. Triton is one and is close to Neptune. The others are the four outermost satellites of Jupiter and the outermost satellite of Saturn. Many believe these satellites to be captured asteroids and not permanent members of the system in question. This topic will be discussed at length in Chapter 6. Whatever their subsequent behaviour, the Voyager mission has given us much information about the very individual compositions of the satellites which may lead to a better understanding of their history and the history of the Solar System as a whole. A general description of these bodies can be found in the Cambridge Atlas of Astronomy (1985).

There are many near commensurabilities in mean motions (average orbital angular velocities) present in the Solar System. By a near commensurability we mean that the ratio of two mean motions may be closely approximated by a simple vulgar fraction. Some commensurabilities are important for maintaining stability. Some examples are Neptune-Pluto (3:2), Titan-Hyperion (4:3), and Enceladus-Dione (2:1). Other commensurabilities such as Jupiter-Saturn (5:2) and Uranus-Neptune (2:1) do not seem to be critical for maintaining stability. Roy and Ovenden (1954, 1955) have shown that there are significantly more occurrences of near commensurabilities than expected by chance. This implies that these resonant states are preferred, and may be selected by a dynamical

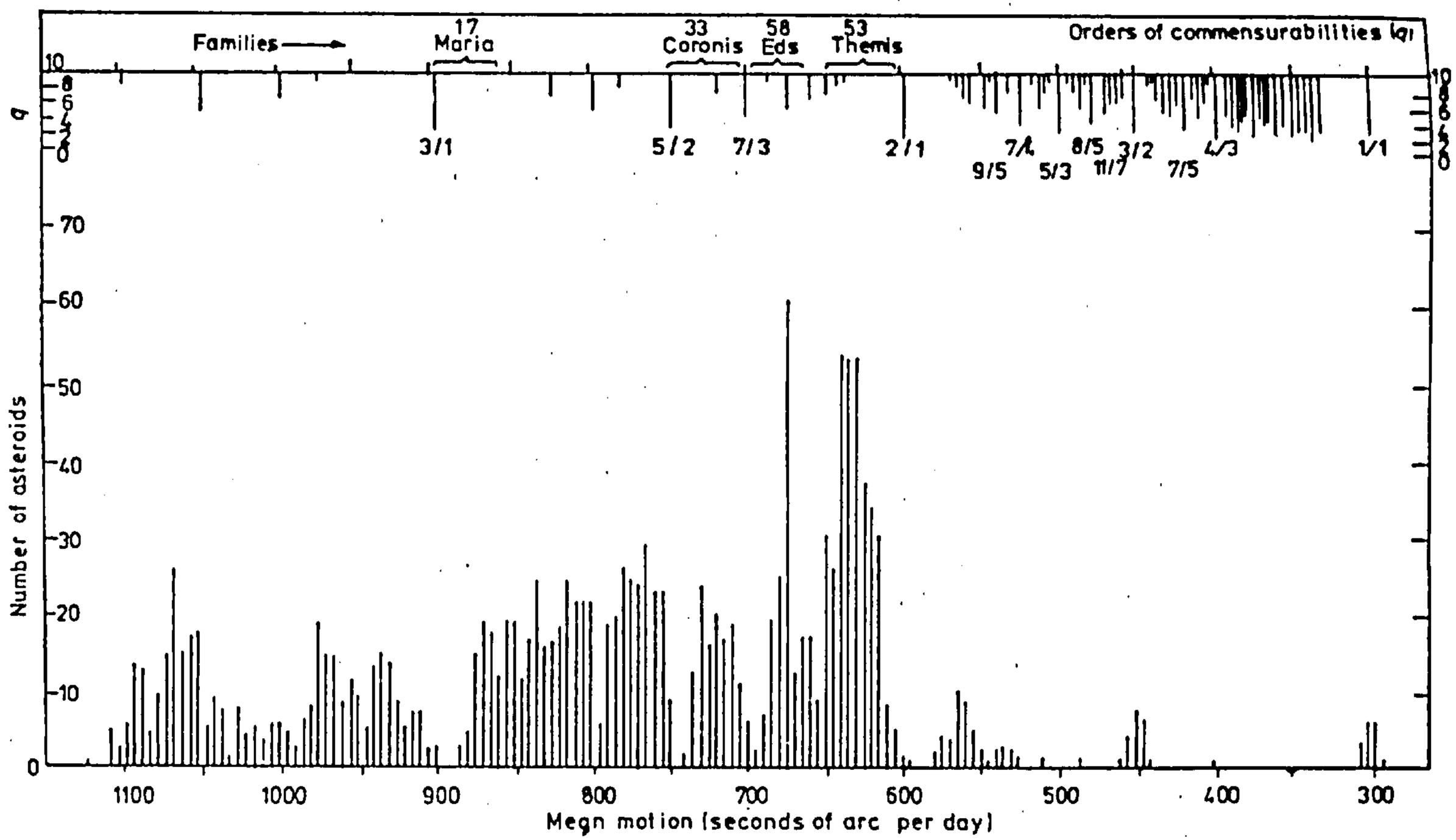
1.1

mechanism, for example, tidal interactions (Goldreich, 1965).

The asteroids are seen in a variety of orbits. Most are in near circular orbits between the orbits of Mars and Jupiter (Figure 1.1), but other families exist which have higher eccentricities and cross planetary orbits, including the Earth's. Examples of the latter type are the Apollo and Amor objects and the Hidalgo group. On examining the main asteroid belt, gaps can be observed at certain distances from the Sun where there are relatively few asteroids. At other distances there is a greater concentration of bodies than average. Figure 1.2 shows the distribution of asteroids with respect to their mean motions. The so-called Kirkwood gaps are clearly seen. Along the top of the figure are given the orders of commensurabilities between Jupiter and the asteroids at the given mean motion. It becomes apparent that the most prominent gaps occur at commensurabilities (3:1), (5:2) and (7:3). The main belt is cut off sharply at the (2:1) commensurability, (called the Hecuba Gap). There are two further clusters of asteroids. These are the Hilda and Trojan groups associated with the (3:2) and (1:1) commensurabilities respectively. The Trojans will be referred to again in connection with the restricted three-body problem, (Section 1.2). At first sight, it might appear that these gaps are regions of unstable motion. However it was suggested by Brouwer (1963) and Message (1966) that the associated commensurabilities are stable and that asteroids are in orbits that oscillate about a mean position given by the gap. Therefore more asteroids would be observed to bound the gap than be present in it.

Ring structure exists on a smaller scale around at least three of

Figure 1.2: Distribution of asteroids with their mean motions about the Sun, in seconds of arc per day. Mean motions which are commensurable with that of Jupiter are indicated (from Brouwer 1963).

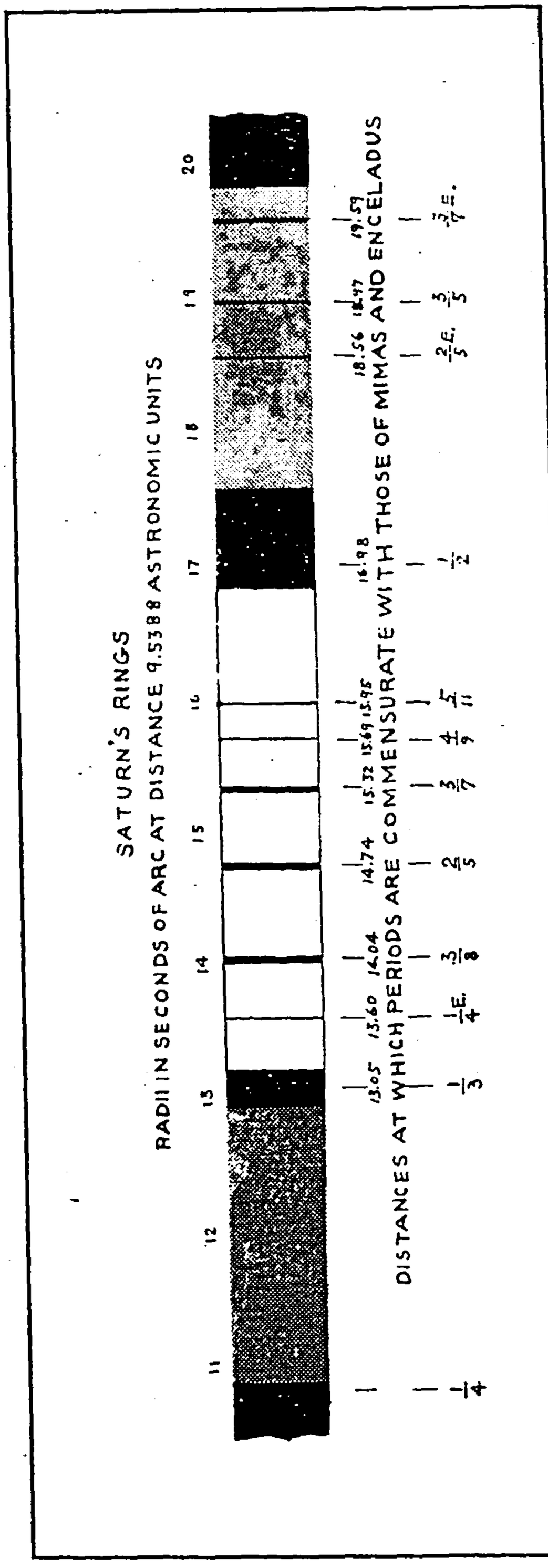


the gas giants. The most noticeable rings are around Saturn. These rings of small particles contain gaps similar to the Kirkwood Gaps which are caused by resonances with nearby satellites. Figure 1.3 shows some of the principal gaps and associated commensurabilities. The largest gap is Cassini's division between the A and B rings, which arises from commensurabilities with Mimas, Enceladus and Tethys in the ratios (2:1), (3:1) and (4:1) respectively. The boundary between the B and C rings is called Encke's division and occurs at a distance which allows for a (3:1) commensurability with Mimas. Images from the Voyager mission have shown that the rings are very finely structured, reflecting many different resonances at work. Small satellites have been found that graze the rings and must play an important part in deciding their evolution.

Comets and meteors are small bodies in highly eccentric, inclined orbits. The most popular explanation for the origin of the comets, given by Oort, is that many millions of them form a shell around the Solar System. Through perturbations by nearby stars, some are sent into the planetary region where subsequent perturbations by Jupiter and the other gas giants, render them in orbits that are either hyperbolic or have semi-major axes comparable with the planetary orbits. As an example, Brook's Comet (1889 V) had a period of revolution 29.2 years, its orbit lying outside Jupiter. On July 20 1886 after a close encounter with Jupiter, its period became 7.10 years, its orbit inside Jupiter's.

Meteors are much smaller silicate bodies that occur in streams whose orbital characteristics are similar to the comets. It may be expected that these streams are also prone to disturbances from the

Figure 1.3


SATURN'S RINGS
RADII IN SECONDS OF ARC AT DISTANCE 9.5388 ASTRONOMIC UNITS
DISTANCES AT WHICH PERIODS ARE COMMENSURATE WITH THOSE OF MIMAS AND ENCELADUS
20
19
18
17
16
15
14
13
12
11
13.05
13.60
14.04
14.74
15.32
15.52
15.69
15.95
16.98
18.56
18.97
19.59
1/4
1/3
3/8
2/5
3/7
1/2
3/5
2/3
4/7
5/7
5/9
4/9
1/E.
2/E.
3/E.

1.1

planets. They cannot however be observed unless they fall through the Earth's atmosphere. It is suspected that meteors originate at the same time as comets or are remnants of disrupted comets. It is hoped that the rendezvous between Halley's Comet and the Giotto Probe this year may shed some light on these questions.

With the possible exception of the comets in Oort's Cloud, all the bodies of the Solar System are effectively isolated from other external influences such as nearby stars or the galactic bulge. Tidal and relativistic effects within the Solar System are more important.

Given that the Solar System is isolated, we may ask various questions concerning its past and future evolution. How old are the planets and the satellites? How stable are the planets' orbits against their mutual gravitational disturbances? Are the satellites in stable orbits or will they be disrupted by dynamical mechanisms such as tidal effects? Given that most bodies orbit in the same direction, how have the retrograde satellites evolved?

Cratering on the Moon and other satellites indicates that in the past there were many collisions with small bodies. The rings are presumably formed by satellites which have been tidally disrupted. There can be little doubt that some bodies are unstable. The reverse question of whether any are stable for all time is a more difficult question to answer.

Records exist from Babylonia around 500 B.C. which describe the motions of Mercury, Venus, Mars, Jupiter and Saturn. They indicate that the planets' orbits differ very little from those followed at the present day. By studying megalithic observatories, we can see that around 3000 B.C. the Moon was moving in the orbit predicted by modern

1.1

lunar theory. These planets have been observed for hundreds of revolutions so drastic changes seem unlikely. The largest satellites have been observed for less than four hundred years. However in that time, they too have performed many thousands of revolutions. A few are observed to be spiralling in towards their respective planets but the majority show little change in the general size and shape of their orbits. Only in the case of the most recently discovered planets; Uranus, Neptune and Pluto, are we short of direct observational evidence. Since their discoveries, Uranus has undergone 2.4 orbits; Neptune, one orbit and Pluto 1/5 of an orbit. We must therefore resort to numerical integration experiments to augment our meagre knowledge of their evolution.

From geological and astrophysical evidence, the Solar System is estimated as being around 4000 million years old. Our observational information over 3000 years is no more than a snapshot when confronted with such timescales. No one should therefore underestimate the difficulties in predicting the long term evolution of the Solar System by extrapolating from these closely grouped data. Thus the questions concerning stability in the Solar System have yet to be answered conclusively.

Having asked the question, "Is the Solar System stable?", we must define what we mean by stable. It is probably still true that the number of workers in the field exceeds the number of definitions of stability. However there are sufficiently many definitions used that extreme care must be taken to specify exactly the conditions for stability. A particular system may be stable according to one definition, yet unstable for another.

1.1

As an example, consider the motion of a planet about the Sun in the absence of any other perturbing bodies. This is the classical two-body problem which yields an analytic solution for the planet's motion. It describes a fixed ellipse, characterised by six constants including the semi-major axis a_1. The time-dependent variable is the true anomaly f_1. If the planet is displaced by a small amount, it will describe a different ellipse with a different semi-major axis a_2, and true anomaly f_2. We examine two definitions of stability. The first by Liapunov asks, "if the two orbits differ only slightly, at some initial time, will the two bodies remain close for all time?" The answer is "no", since by changing the semi-major axis, the orbital period is also changed and the bodies will drift apart with time, the maximum separation being the major axis of their orbits. The second definition by Poincaré asks, "if the two orbits differ only slightly at some initial time, will the differences in size, shape and orientation remain small for all time?" The answer in this case is "yes".

Birkhoff (1927) found it appropriate to define time-dependent stability when a body remains in a specified neighbourhood for a specified time. (If the time is infinite we have complete stability).

Hagihara (1957) asks the converse question, "After what length of time will a system deviate from its initial conditions by a previously specified amount?"

There is the possibility that a system may or may not be stable depending on the choice of initial conditions. There are therefore many choices of definition, each appropriate in its own way. Comparisons have been made by many authors, eg: Jefferys and Szebehely (1978), Szebehely

1.1

(1985). The rest of this chapter is devoted to looking at some of the more important definitions of stability as well as mentioning any relevant applications.

1.2 The Restricted Three-Body Problem and Hill Stability

Much of the work of this thesis will be concerned with the study of the general three-body problem. Given three bodies of finite mass and given their initial positions and velocities, the object is to derive their positions within their mutual gravitational field at any future time. In order to solve the nine second order differential equations, it is necessary to have 18 integrals of the motion. Unfortunately only 10 exist, thus the problem does not admit an analytical solution. Even by eliminating the time as independent variable and carrying out the so-called 'elimination of the nodes' procedure, the problem is still of sixth order.

The complexity of the general problem has prompted many workers to consider a simpler model where one of the three bodies has an infinitesimally small mass. In this case the two massive bodies describe circular orbits, unperturbed by the particle. This so-called circular restricted three-body problem requires us to derive the orbit of the third body in the gravitational field of the other two (a reduction from 18 unknowns to 6). The problem may be simplified further by considering only coplanar orbits, in which case, there are only four unknowns, namely the position and velocity of the particle in two dimensions.

In order to make this simplification, we have been forced to discard the ten integrals. However, Jacobi (1836) derived another integral of the motion (Jacobi's integral) which is related to the total

1.2

energy of the system. This integral was used by Hill (1878), to study the stability of the particle's orbit. Hill stability is also a useful concept for the general three-body problem. Therefore, by way of introducing the work of subsequent chapters, we employ it in the circular restricted three-body problem, (Danby, 1962; Roy, 1978). Denote the two massive primaries by P_1 and P_2 (see Figure 1.4).

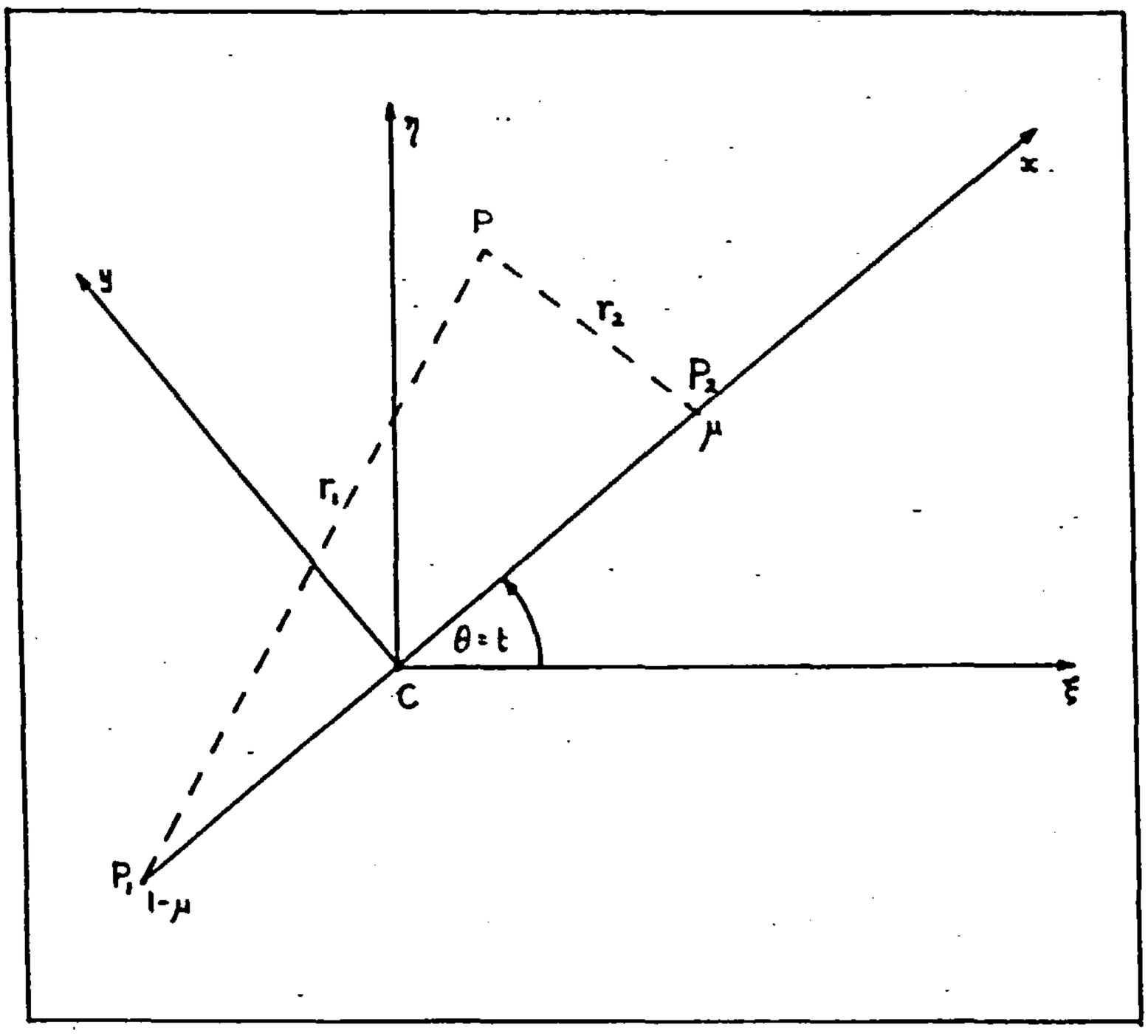


Figure 1.4: The Circular Restricted Three-Body Problem. The two massive bodies are P_1, P_2 and the particle is P.

We take the unit of mass to be the total mass of the primaries, $m_1 + m_2$. Hence we denote the mass of P_1 and P_2 by $1-\mu$ and μ respectively. Without loss of generality, assume that P_2 is less massive.

1.2

Hence $0 < \mu \leq \frac{1}{2}$. Let P_1 and P_2 describe circular orbits about their common mass-centre C. Hence their separation is a constant which is taken to be the unit of length. Kepler's third law states that

$$n^2 a^3 = G(m_1 + m_2) \qquad (1)$$

where a is the separation and n the mean motion of the primaries. (In the case of circular orbits, n is equal to the angular velocity). If we choose the unit of time such that the gravitational constant G is unity, then the angular velocity is also unity.

The equations of motion of a particle P, placed in the gravitational field of P_1 and P_2, are now required. In a non-rotating coordinate system, with its origin C; P_1, P_2, P have coordinates (ξ_1,η_1,ζ_1), (ξ_2,η_2,ζ_2), (ξ,η,ζ) respectively. The equations of motion are

$$\ddot{\xi} = (1-\mu)\frac{\xi_1-\xi}{r_1^3} + \mu\frac{\xi_2-\xi}{r_2^3} \qquad (2)$$

$$\ddot{\eta} = (1-\mu)\frac{\eta_1-\eta}{r_1^3} + \mu\frac{\eta_2-\eta}{r_2^3} \qquad (3)$$

$$\ddot{\zeta} = (1-\mu)\frac{\zeta_1-\zeta}{r_1^3} + \mu\frac{\zeta_2-\zeta}{r_2^3} \qquad (4)$$

where '.' refers to differentiation with respect to time and

$$r_1 = [(\xi_1-\xi)^2 + (\eta_1-\eta)^2 + (\zeta_1-\zeta)^2\ ]^{\frac{1}{2}}$$
$$r_2 = [(\xi_2-\xi)^2 + (\eta_2-\eta)^2 + (\zeta_2-\zeta)^2\ ]^{\frac{1}{2}} \qquad (5)$$

Since P_1 and P_2 are not disturbed by P, it is assumed that (ξ_1,η_1,ζ_1) and (ξ_2,η_2,ζ_2) are known functions of time. Without loss of generality we orientate the coordinate axes so that the motion of the primaries

1.2

is wholly in the $\xi\eta$-plane, i.e. $\zeta_1 = \zeta_2 = 0$ for all time.

We shall be more interested in the motion of P <u>relative</u> to P_1 and P_2. We therefore adopt a new rotating coordinate system x,y,z with the origin still at C and the z-axis coinciding with the ζ-axis. However the system rotates with angular velocity equal to unity, such that the primaries always lie on the x-axis. In the new system, P_1 and P_2 have coordinates $(-\mu, 0, 0)$ and $(1-\mu, 0, 0)$ respectively.

Let Θ be the angle $\xi\hat{C}x$ at a given time t. Then the new coordinates of P (x,y,z) are given by

$$\begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} = \begin{bmatrix} \cos\Theta & -\sin\Theta & 0 \\ \sin\Theta & \cos\Theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (6)$$

Since the angular velocity is unity, $\Theta = t + \text{constant}$. This constant may be set to zero without loss of generality. Using Matrix Equation (6), we may obtain ξ, η, ζ and their derivatives. On substitution in Equations (2), (3), (4),

$$(\ddot{x}-2\dot{y}-x)\cos t - (\ddot{y}+2\dot{x}-y)\sin t =$$

$$\left[(1-\mu)\frac{x_1 - x}{r_1^3} + \mu\frac{x_2 - x}{r_2^3}\right]\cos t + \left[\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right]\sin t \qquad (7)$$

$$(\ddot{x}-2\dot{y}-x)\sin t + (\ddot{y}+2\dot{x}-y)\cos t =$$

$$\left[(1-\mu)\frac{x_1 - x}{r_1^3} + \mu\frac{x_2 - x}{r_2^3}\right]\sin t - \left[\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right]\cos t \qquad (8)$$

$$\ddot{z} = -\left[\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right] z \qquad (9)$$

1.2

where

$$r_1 = [(x_1 - x)^2 + y^2 + z^2]^{\frac{1}{2}} \qquad (10)$$

$$r_2 = [(x_2 - x)^2 + y^2 + z^2]^{\frac{1}{2}}$$

and $x_1 = -\mu \quad , \quad x_2 = 1-\mu \qquad (11)$

If we multiply Equation (7) by cos t and Equation (8) by sin t and add, then multiply Equation (7) by - sin t and Equation (8) by cos t and add, we obtain two equations which, along with Equation (9) comprise the equations of motion of P in the rotating coordinate frame, viz.

$$\ddot{x} - 2\dot{y} = \frac{\partial U}{\partial x} \qquad (12)$$

$$\ddot{y} + 2\dot{x} = \frac{\partial U}{\partial y} \qquad (13)$$

$$\ddot{z} = \frac{\partial U}{\partial z} \qquad (14)$$

where

$$U = \frac{1}{2}(x^2+y^2) + \frac{1-\mu}{r_1} + \frac{\mu}{r_2} \qquad (15)$$

Multiplying Equations (12), (13), (14) by $\dot{x},\dot{y},\dot{z}$ respectively and adding, we obtain

$$\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} = \frac{\partial U}{\partial x}\dot{x} + \frac{\partial U}{\partial y}\dot{y} + \frac{\partial U}{\partial z}\dot{z} \qquad (16)$$

Since U does not depend on t explicitly, Equation (16) is a perfect differential which we may integrate once to obtain

$$2U - (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = C \qquad (17)$$

where C is a constant of integration known as *Jacobi's integral*.

1.2

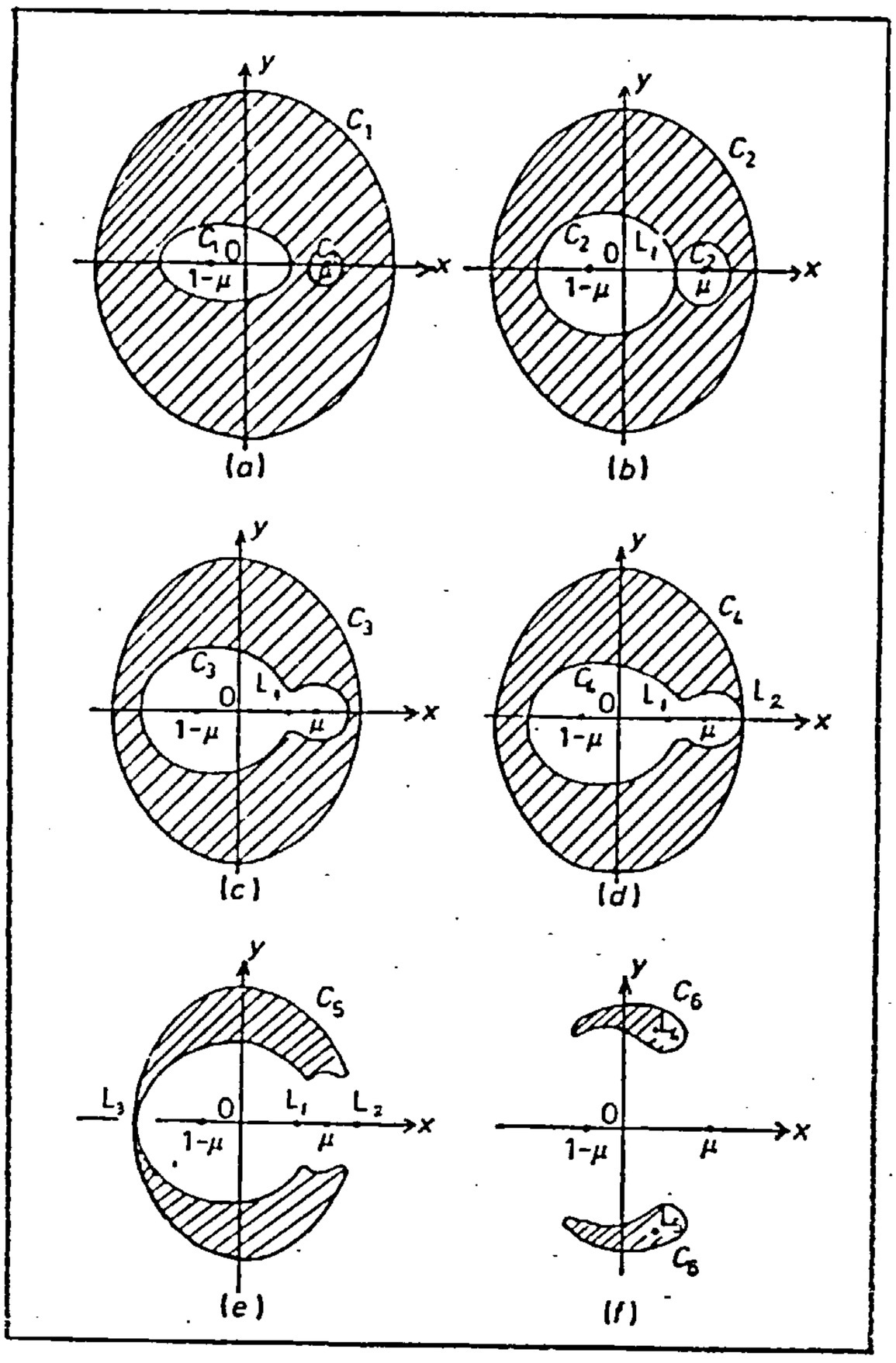
Since $(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)$ is always positive or zero it follows that $2U \geq C$. For a given value of C, there are therefore regions in the coordinate space where we may position the particle ($2U \geq C$) and regions where we may not ($2U < C$, implying an imaginary relative velocity). Clearly the boundary is given when $\dot{x}^2 + \dot{y}^2 + \dot{z}^2$ equals zero, i.e. the velocity of P in the rotating frame is zero. These zero-velocity surfaces are given by

$$x^2 + y^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} = C \qquad (18)$$

We now examine the topology of these surfaces as C varies, although it should be remembered that for any given dynamical system, C is an integral of the motion and hence constant for all time. It is from such an examination that we derive the concept of *Hill Stability*.

Consider the case when the particle is restricted to the (x,y)-plane (coplanar motion). Sets of curves are presented in Figures 1.5 which represent Equation (18) as C is varied. The region of forbidden motion is shaded.

For $C = C_1 >> 1$ (Figure 1.5(a)), there are three disconnected regions of allowed motion; one bounded region around each primary and an unbounded region when $x^2 + y^2$ is sufficiently large. The particle is confined to one of the three regions and is prohibited from moving to the other two. If the particle is confined to one of the bounded regions around a primary, it can never approach the other binary or escape to infinity. This consideration of the boundedness of the particle's orbit is the basis of Hill stability. We can say nothing concerning the character of the orbit within the region. We do not



Figures 1.5. The zero-velocity curves of the coplanar circular restricted three-body problem.

1.2

know how close it may approach the primary which it is orbiting. All we know is that it is constrained to remain close to that primary.

If C is decreased to C_2 (Figure 1.5(b)), the bounded regions are seen to have a common tangent at the point L_1. If C is reduced further to C_3, (Figure 1.5(c)) then the bounded regions coalesce to form a dumb-bell shape. The particle, if confined to this region, is free to wander from one primary to the other (although it is not obliged to do so). It is still unable to escape to infinity from this region.

If C is decreased further to C_4 (Figure 1.5(d)), the two remaining regions of allowed motion have a common tangent at the point L_2. If C is reduced further then these regions coalesce. The particle is now free to move between primaries or wander into the external area.

Once $C = C_5$, (Figure 1.5(e)), a common tangent occurs at the point L_3. Further reductions in C (Figure 1.5(f)), result in there being two disconnected regions of forbidden motion, which shrink eventually to points L_4 and L_5, where $r_1 = r_2 \approx 1$. Sections through the zero-velocity surfaces in the xz- and yz-planes are shown in Figures 1.6 and 1.7 for the same values of the Jacobi Integral as are given in Figures 1.5.

The critical values of the Jacobi Integral are C_2,C_4 and C_5, which determine the connectedness of the various regions of allowed motion. Hill stability can be applied to a variety of dynamical systems.

(a) Consider a massless planet in an inferior orbit about the larger star of a double star system ($\mu \lesssim \frac{1}{2}$). Alternatively, consider a planet in an inferior orbit about the Sun, perturbed by a massive superior planet ($\mu << \frac{1}{2}$). If $C \geq C_2$. then the planet is constrained

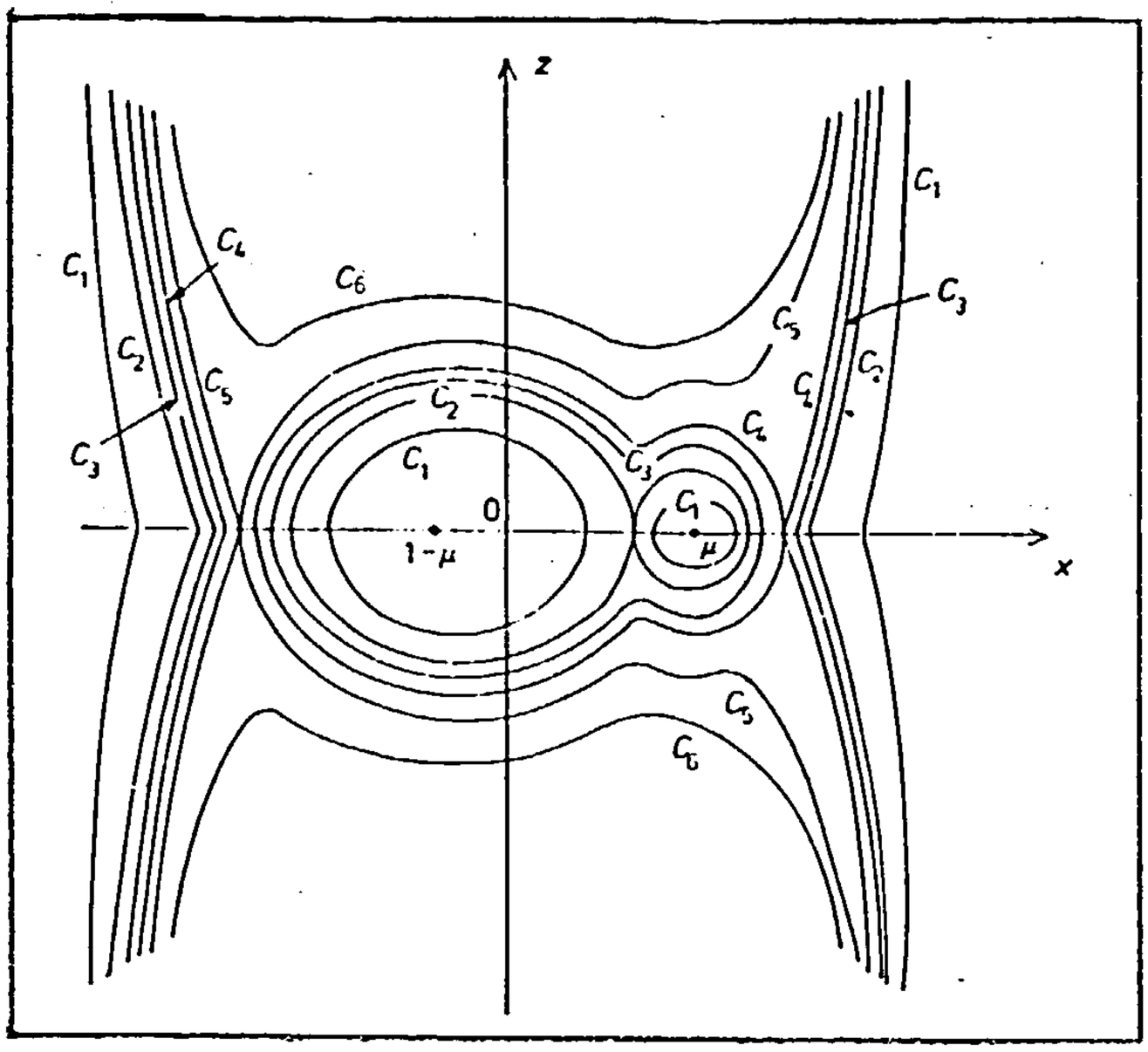


Figure 1.6: Section through the zero-velocity surfaces of the circular restricted three-body problem in the xz-plane.

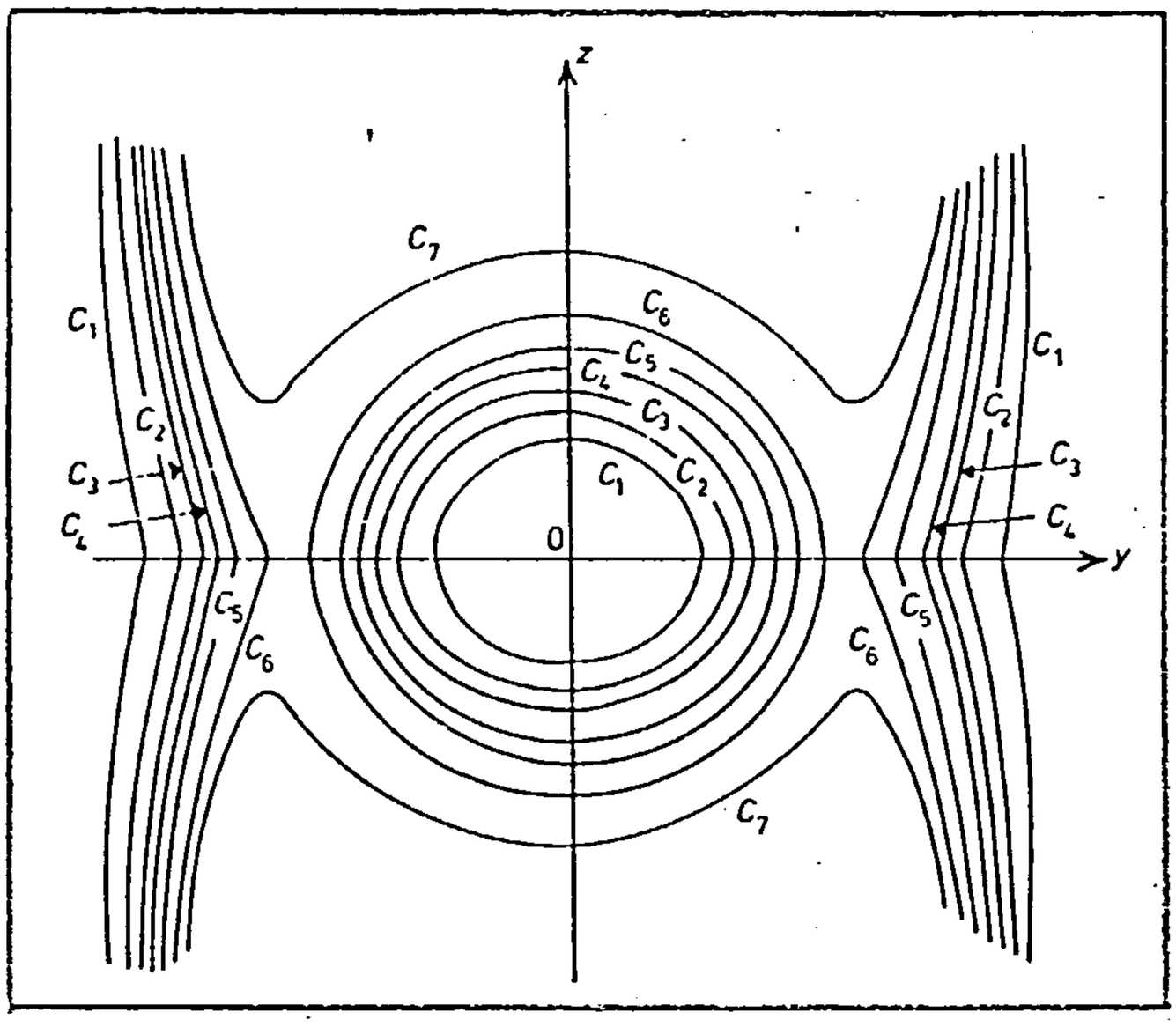


Figure 1.7: Section through the zero-velocity surfaces of the circular restricted three-body problem in the yz-plane.

1.2

to orbit the larger primary. If $C_4 \leq C < C_2$, then it could transfer its orbit to P_2 or alternate between P_1 and P_2. If $C < C_4$, then the planet could escape from the two primaries altogether.

(b) Consider a massless planet in an inferior orbit about the smaller star of a double star system ($\mu \lesssim \frac{1}{2}$), or a satellite orbiting a planet ($\mu << \frac{1}{2}$). In both these cases the argument is identical to (a).

(c) Consider a (massless) body in a superior orbit about a double star system ($\mu \lesssim \frac{1}{2}$) or a Sun-planet system ($\mu << \frac{1}{2}$). The particle cannot approach either primary so long as $C \geq C_4$. In this case it does not matter if the two regions around the primaries are connected.

It is possible to use the initial conditions to calculate C from Equation (17). We may then compare it with the critical values C_2 and C_4 to determine whether the particle is prohibited from approaching one or both primaries.

Critical values of C occur when the particle is at one of the points L_1, L_2, L_3, L_4, L_5, (the most important being L_1 and L_2 when considering Hill stability). These points are double points where tangents to the zero-velocity curves coincide. By definition, this is where the partial derivatives of the function 2U-C vanish. (Recall that 2U=C is the equation of the zero-velocity curves). Hence at these points

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} = \frac{\partial U}{\partial z} = 0 \qquad (19)$$

However, being zero-velocity curves, $\dot{x} = \dot{y} = \dot{z} = 0$. By Equations (12), (13) and (14), $\ddot{x}$, $\ddot{y}$, $\ddot{z}$ are all zero, implying there are no resultant forces on the particle. L_1, L_2, L_3, L_4 and L_5 are therefore

1.2

the familiar Lagrange equilibrium solutions. The L_1 and L_2 points may be found by solving Equations (19) with $y = z = 0$. Hence C_2 and C_4 may be evaluated and the Hill stability of the system determined.

It is possible to study the stability of the particle at these equilibrium points (x_o, y_o, z_o) by displacing it a small amount to a new position $(x_o + x', y_o + y', z_o + z')$. If the particle oscillates about this point, it is defined to be in a stable position. If the particle departs from the neighbourhood of the point, it is unstable. The equations of motion, (12), (13), (14) may be expanded as Taylor series in powers of x', y', and z'. Because the displacement is small, we retain only the linear terms. The general solution of these linearised differential equations may be expressed in the form

$$x' = \sum_{i=1}^{6} \alpha_i \exp(\lambda_i t)$$

$$y' = \sum_{i=1}^{6} \beta_i \exp(\lambda_i t)$$

$$z' = \sum_{i=1}^{6} \gamma_i \exp(\lambda_i t)$$

where α_i are constants of integration, β_i and γ_i depending on α_i. The λ_i are obtained by substituting $x' = \alpha e^{\lambda t}$, $y' = \beta e^{\lambda t}$, $z' = \gamma e^{\lambda t}$ into the linearised differential equations. By eliminating α, β, γ among the three equations we are left with a polynomial equation in λ, with highest power λ^6. The solution of this equation provides at most six distinct values of λ. If all of them are purely imaginary then x', y', z' are purely oscillatory with time and hence give stable solutions.

When this linear stability criterion is applied to the five Lagrange

1.2

points it is found that L_1, L_2 and L_3 are always unstable while L_4 and L_5 are stable provided $\mu < 0.0385$. A practical example of this stability can be found in the Trojan asteroids which exist at the L_4 and L_5 points of the Sun and Jupiter.

In recent years, there has been considerable interest in the circular restricted three-body problem when non-gravitational forces are taken into account. Particular attention has been paid to the existence and stability of equilibrium points when radiation pressure from one or both primaries is considered, (Schuerman, 1980; Mignard, 1982; Mignard and Henon, 1984; Simmons et al; 1985). The effect on a charged particle in the presence of two revolving dipoles is an allied problem which has been studied by Goudas and Petsagourakis (1985).

Hill stability has been applied to many real systems. Szebehely and McKenzie have applied it to the cases Sun-Jupiter-Saturn (1977a) and Earth-Moon-Sun (1977b) and shown them to be stable. Hagihara (1952) applied the Jacobi integral to the other natural satellites of the Solar System. The zero-velocity surfaces have been used extensively to study matter transfer between binary star components (Roy, 1978; Boyle, 1984).

It must be remembered however that the results derived in these cases are only as valid as the model they are derived from, namely the circular restricted three-body problem. In general the orbits will have non-zero eccentricities and the "particle" may be rather massive (particularly in the case when Saturn is assumed to be massless beside the Sun and Jupiter). Using the theory that will be described in Chapter 3, Szebehely and McKenzie have shown that relaxing the conditions on the eccentricities and masses produces profound changes on the Hill-type stability. Indeed there is no longer any guarantee

1.2

that the Earth-Moon-Sun system is stable.

When the primary bodies are allowed to describe elliptic orbits, there is no Jacobi integral to which we can apply Hill stability. It is possible to derive formal expressions for the angular momentum and Jacobi integrals (Ovenden and Roy, 1961; Sarris, 1982), but they are explicitly time-dependent. The zero-velocity surfaces vary with time which implies that Hill stability cannot be guaranteed at future epochs.

Because of these problems, Hill stability in the restricted three-body problem is of limited use when applied to the Solar System, except when dealing with circular orbits and very small satellites. A more useful generalisation of Hill stability in the general three-body problem is presented in Chapter 3.

1.3 Periodic Orbits

Periodic orbits have played an important part in the study of dynamical systems this century. They may be defined most generally as follows. A system comprising n bodies can be described at any instant by a point in the 6n-dimensional phase space defined by the 6n spatial coordinates and velocities of the bodies. Denote this position by vector $\underset{\sim}{s}$ which is a function of time t. As t varies, so does $\underset{\sim}{s}$ and the system describes a trajectory in the phase space corresponding to an orbit in the 3-dimensional coordinate space. A *periodic orbit* is defined to be an orbit where

$$\underset{\sim}{s}(t) = \underset{\sim}{s}(t + T) \qquad (20)$$

for any t and some fixed value T. If T is the smallest value such that

1.3

Equation (20) holds for all t, then it is called the *period* of the orbit.

Since Poincaré (1895) applied periodic orbits to the restricted three-body problem, many workers have made comprehensive studies of their properties. In recent years with the advent of faster computers, workers have been able to study periodic orbits in the general three-body problem as well as other conservative and dissipative Hamiltonian systems. This has allowed studies into a variety of problems such as galactic rotation (Contopoulos, 1983a). Although the concepts described in this section apply to general dynamical systems, we shall mostly refer to the restricted three-body problem.

Roy and Ovenden (1955) describe their mirror theorem as follows: if n point masses are acted on by their mutual gravitational forces only, and at a certain epoch, every radius vector from the centre of mass of the system is perpendicular to every velocity vector, then the orbit of each mass after that epoch is a mirror image of its orbit prior to that epoch. Such a configuration of radius and velocity vectors is called a *mirror configuration*. As a result of this theorem, they point out that any system which undergoes two mirror configurations must be a symmetric periodic orbit.

There are many reasons why periodic orbits are useful in practical situations:

(i) The conjecture of Poincaré states that for any bounded solution to the equations of motion of a dynamical system, it is possible at any time to find a periodic solution which is arbitrarily close to the original, (although the same periodic solution may not remain close

1.3

for all time). For this reason periodic solutions have applications as reference orbits, for example in Encke-type numerical algorithms (Section 1.6).

(ii) Many real dynamical systems are found to exhibit resonances, for which periodic orbits are directly relevant to stability.

(iii) Periodic orbits may be found and classified by analytical and numerical techniques. In this way it is possible to find all the periodic solutions to a given problem. The stability of more general solutions may be found by studying the stability of nearby periodic orbits.

It is possible to classify periodic orbits according to their initial conditions. A group of periodic orbits whose initial conditions vary continuously one to another is called a *family*. Due to this continuity, it has proved fruitful to search for distinct families using numerical techniques applied to an analytical continuation theory. Markellos (1974a) describes one such method for the restricted three-body problem, replacing one initial velocity component by Jacobi's integral to act as an independent variable.

A linear stability analysis, similar to that described in Section 1.2 may be applied to periodic orbits. Let $\underset{\sim}{s}_o$ correspond to an initial state of the orbit and $\underset{\sim}{s}$ correspond to a state which is only slightly displaced from $\underset{\sim}{s}_o$. Define the variational matrix

$$\Delta(\underset{\sim}{s}_o, t) = \frac{\partial(s_i)}{\partial(s_{oj})} \qquad (21)$$

where s_i, s_{oj} refer to components of $\underset{\sim}{s}$, $\underset{\sim}{s}_o$ respectively. The condition for stability of the periodic orbit at $\underset{\sim}{s}_o$ is that all the

1.3

eigenvalues of $\Delta(s_o, t)$ have modulus equal to unity. For the restricted three-body problem there are four eigenvalues, two of which are always equal to unity. The other two have product unity and may be represented by $\exp(\lambda T)$ and $\exp(-\lambda T)$, where T is the period of the orbit. λ and $-\lambda$ are called Poincaré's characteristic exponents and must be purely imaginary for stability.

Another important concept for stability analysis is the surface of section. In the circular coplanar restricted three-body problem, the position and velocity of the particle relative to the primaries, $(x,y,\dot{x},\dot{y})$ define its orbit (see Section 1.2). By choosing a particular value of Jacobi's integral C, we may express $\dot{y}$ in terms of $x,y,\dot{x}$. Hence the orbit in phase space is given by the locus of points $(x,y,\dot{x}, C')$. A particular value of C is chosen. We wish to examine the particle's orbit whenever y is a particular value y^* and $x > 0$. We therefore define a mapping M^n which maps a point in the orbit $(x_o,y^*,\dot{x}_o,C)$ to the point in the orbit $(x,y^*,\dot{x},C)$ which is reached as the orbit crosses the line $y = y^*$ $(x > 0)$ for the n^{th} time. For example, if $y^* = 0$ and $n = 1$, we are examining $(x,\dot{x})$ everytime the orbit crosses the positive x-axis. In this case, if the orbit closes after one synodic period then it is characterised by one invariant point on the plane $(x,\dot{x})$, $x > 0$. This plane is called Poincaré's surface of section and is only one of many examples of surfaces of section. If the orbit closes after two synodic periods then it will be characterised by two invariant points on the surface of section defined by M^1 or by one invariant point on the surface of section defined by M^2.

Figure 1.8 gives an example of a surface of section for the case

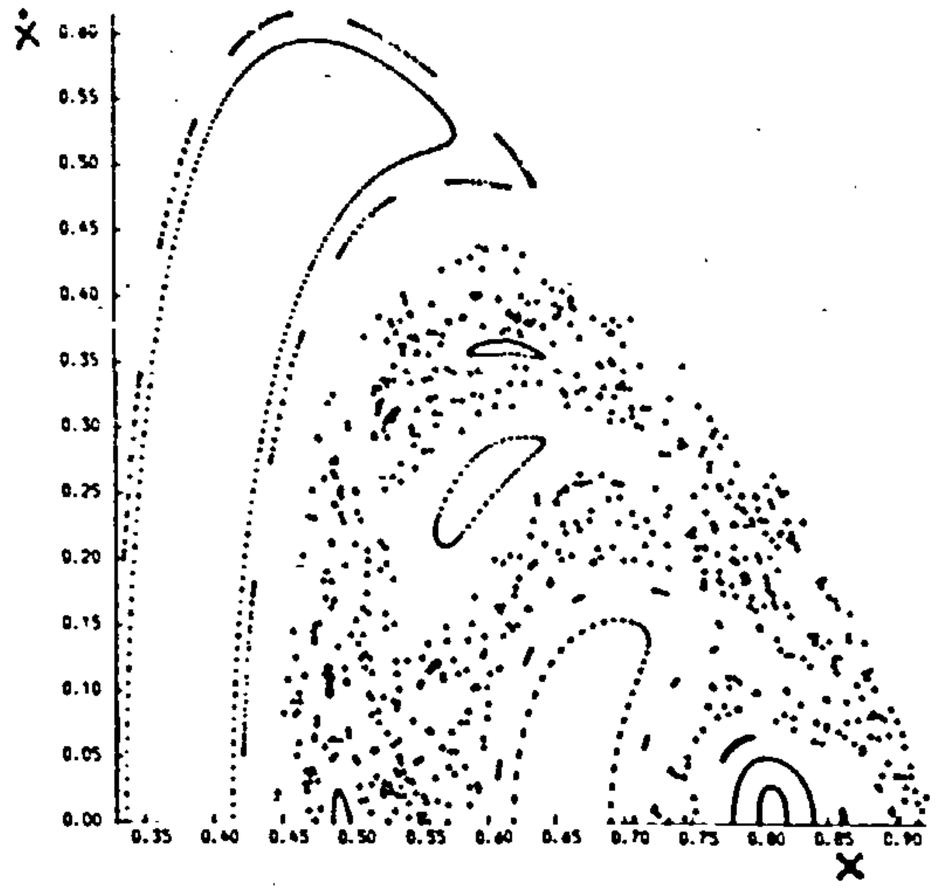


Figure 1.8: Surface of section y=0, x>0 on a fixed value C = -1.52 of the Jacobi integral in the Sun-Jupiter-Asteroid restricted circular problem. Each point gives the x, ẋ values when the orbit crosses the y=0 axis with x>0; the x axis is aligned with the Sun-Jupiter vector and Jupiter is at x=1-μ. The chaotic region (crosses) has been obtained by following the same orbit for a few hundred synodic periods.
(Reproduced from Milani & Nobili, 1985).

of Sun-Jupiter-Asteroid. A symmetric periodic orbit which closes after one synodic period is represented by a single point on the x-axis. If the mapping M^1 implies the point is elliptic, then the corresponding periodic orbit is stable. This is observed when a starting point near an elliptic point is mapped onto a closed (invariant) curve around the elliptic point. As the starting point is moved further away from the elliptic point the invariant curve breaks up into a series of islands which gradually shrink into a set of invariant points, corresponding to a periodic orbit of longer period mT, where T is the period of the elliptic point and m is the number of island points. Each stable island point may in turn have invariant curves around it and so on.

There are also regions where stable periodic orbits do not exist. In these *chaotic* regions a point on the surface of section is mapped to other points within a certain region but in a random manner (Figure 1.8). It is clearly seen from this figure that the surface of section may be divided into regions of chaotic (unstable) motion and regions where invariant curves may be found around stable elliptic points.

Hénon (1966) describes a method for finding the extent of the "region of curves" around elliptic points, requiring numerical integration of the variational equations. There is a further problem in that

the boundary between stability and instability is not clearly defined. Markellos (1974b) points out that, due to the set of rational numbers being dense in the set of real numbers, chains of islands lie arbitrarily close to any invariant curves. This guarantees the existence of invariant points of the mapping M^n where n may be as large as we please. In other words, regions of invariant curves correspond to regions of chains of islands. He describes an alternative method of finding these regions by examining regions where branching (bifurcations) of periodic orbits occur, resulting in new families of higher period. (Contopoulos and Pinotsis (1984) discuss infinite sequences of bifurcations in the restricted three-body problem).

The circular restricted three-body problem is an example of a dynamical system of two degrees of freedom. When the dynamical system has three (or more) degrees of freedom, the difficulties in using periodic orbits are greater. New concepts such as inverse bifurcations and complex instability are introduced and surfaces of section occur in three (or more) dimensions. (See Contopoùlos, 1983 b,c).

Because the stability analysis described here requires the system of equations to be linearised we must neglect the higher order terms. If the linearised system is unstable, this implies that the original system is also unstable. However a stable linearised system does *not* imply that the original system is stable. This is a considerable drawback when trying to consider long term stability except where resonances are involved.

Before leaving this section it is worth noting one result by Hénon (1970). In the circular restricted three-body problem, on examining the zero velocity curves, a Hill stability examination implies that retrograde particle orbits should be stable for a smaller range of initial conditions than direct orbits. By examining regions of invariant curves, Hénon concluded that retrograde orbits were stable for a larger range of initial conditions than direct orbits. Chapters 3-6 of the work presented here are concerned with verifying a similar result for the general three-body problem.

1.4 Commensurabilities

In Section 1.1, examples within the Solar System of commensurabilities in mean motion were given. When two bodies orbit another more massive body, they are said to be commensurable if the ratio of their mean motions approximates to a rational number m/n where m and n are integers. Because rational numbers are dense in the set of real numbers, any two bodies can be defined as commensurable to arbitrary precision by taking m and n large enough. In general however when we refer to a commensurable system we may take it that m and n are reasonably small ($\lesssim 10$).

The importance of commensurabilities may be judged in the light of general perturbation theory (Plummer, 1918). The motion of a body in the presence of other bodies may be derived from the Lagrange planetary equations which make reference to a quantity called the *disturbing function* (see Section 1.5). This function is generally made up of an infinite series of periodic terms, each term involving a linear combination

Text cut off in original

1.4

of the fundamental frequencies of the system. In order to solve the planetary equations it is necessary to integrate these terms, whereupon the linear combinations of frequencies appear as divisors in each term. A commensurable system may render some of these divisors very small, hence magnifying the effect of these terms on the motion of the bodies. It may be shown that the principal effect of a near-commensurability will be observed in perturbations of the mean longitude. (eg. Roy, 1979).

As was mentioned in Section 1.1, the number of commensurabilities observed in the Solar System is significantly greater than expected by chance (Roy and Ovenden, 1954, 1955). Commensurabilities often manifest themselves in the appearance of stable resonant behaviour, where the conjunction line of two satellites orbiting a planet librates about a specific direction, often the apse line of one of the satellites. A quantity Θ may be defined which is a linear combination of mean longitudes λ and longitudes of pericentre $\tilde{\omega}$. One of the best known examples is that of Neptune and Pluto where $\Theta = 3\lambda_P - 2\lambda_N - \tilde{\omega}_P$ oscillates about 180^o with an amplitude of $\sim 80^o$ and a period of 19440 years (Cohen et al., 1967). In the case of the Saturnian satellites Enceladus and Dione, $\Theta = 2\lambda_D - \lambda_E - \tilde{\omega}_E$ oscillates about 0^o with an amplitude of 1.5^o.

Goldreich (1965) proposed that tidal forces between planets and satellites could stabilise such resonant systems. Due to tidal friction, angular momentum from the spinning planet is transferred to the two satellites in the form of orbital angular momentum causing their semi-major axes to increase. The inner satellite spins outwards faster. In this way the system may evolve from a non-commensurable state to a commensurabl

1.4

state. Having reached that state, the angular momentum is shared between the satellites in such a manner as to maintain the commensurability. This assumes that the gravitational interaction between the satellites is strong enough. The method for stabilising the resonance may be illustrated by the following example (see Figures 1.9 a,b).

Consider two satellites orbiting planet P: S_i the inner and S_o the outer with periods T_i and T_o respectively, ($T_i < T_o$). Assume that tidal forces have caused the system to evolve into a resonant state. Suppose that a conjunction of the satellites occurs just prior to S_o being at aphelion. After a further synodic period, the two satellites will be at conjunction once more. However angular momentum will have been transferred from S_o to S_i, due to the asymmetry of the configuration. This will result in T_o shortening and T_i lengthening. Hence the conjunction line will have drifted closer to the apse line of S_o. This process will continue over successive synodic periods until the line of conjunction crosses the apse line. When this happens the process is reversed. S_i gives up angular momentum to S_o, T_o increases, T_i decreases and the line of conjunction begins to drift back towards the apse line once more. In this manner a critical argument may be observed.

Another way in which commensurable behaviour is manifested is the *Laplacian* resonance between three satellites. The most famous example is that of the Galilean satellites Io, Europa and Ganymede where $\lambda_I - 3\lambda_E + 2\lambda_G = 180^o$. This system has been studied by many workers. For example, the stability of periodic orbits in the vicinity of the real solution has been examined, (Wiesel, 1980). A less well known example of a Laplacian resonance can be found in the three Uranian

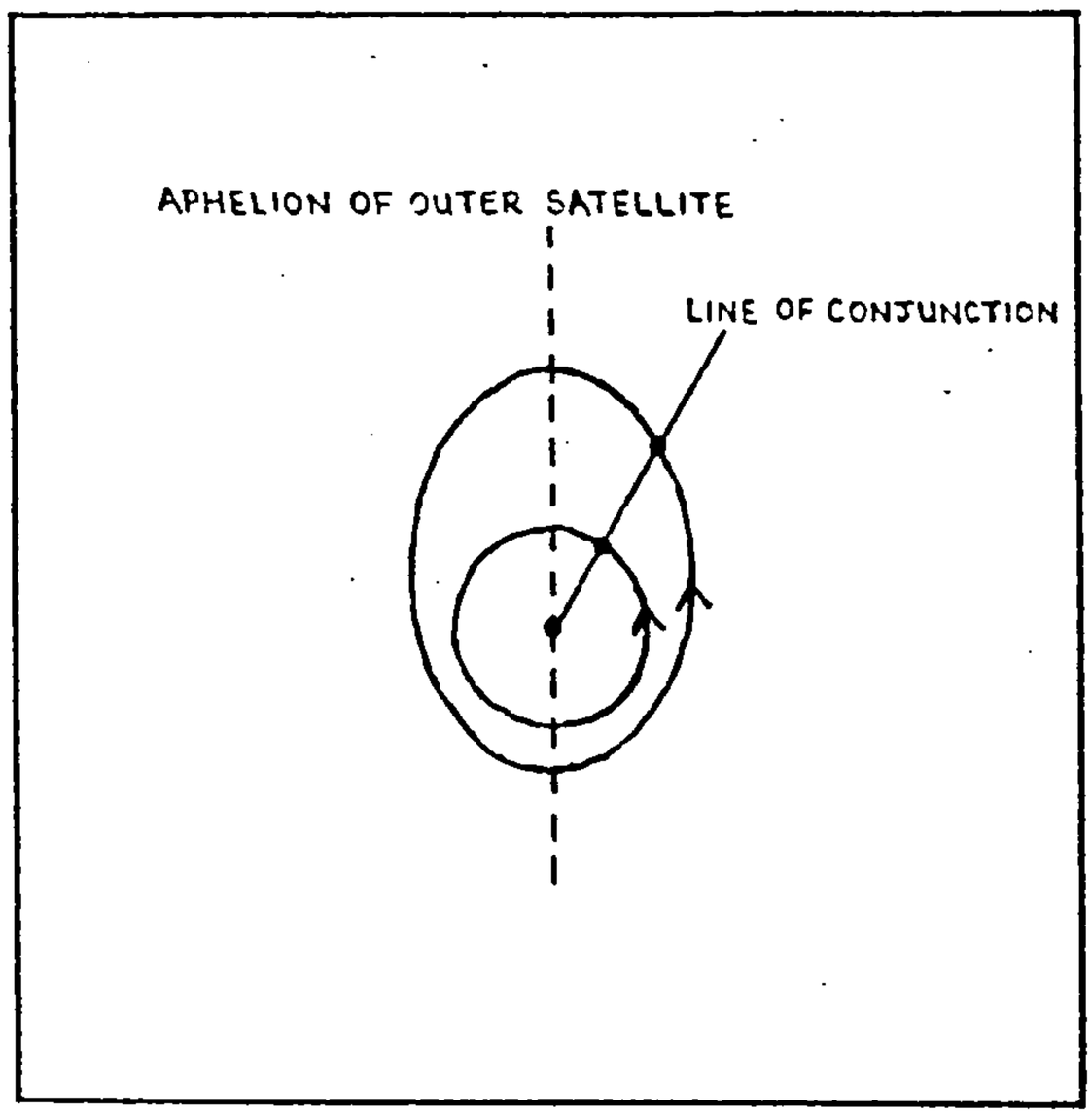


Figure 1.9(a): If two satellites are in resonance, and their conjunction occurs before reaching the aphelion of the outer satellite, angular momentum is transferred from the outer to the inner body.

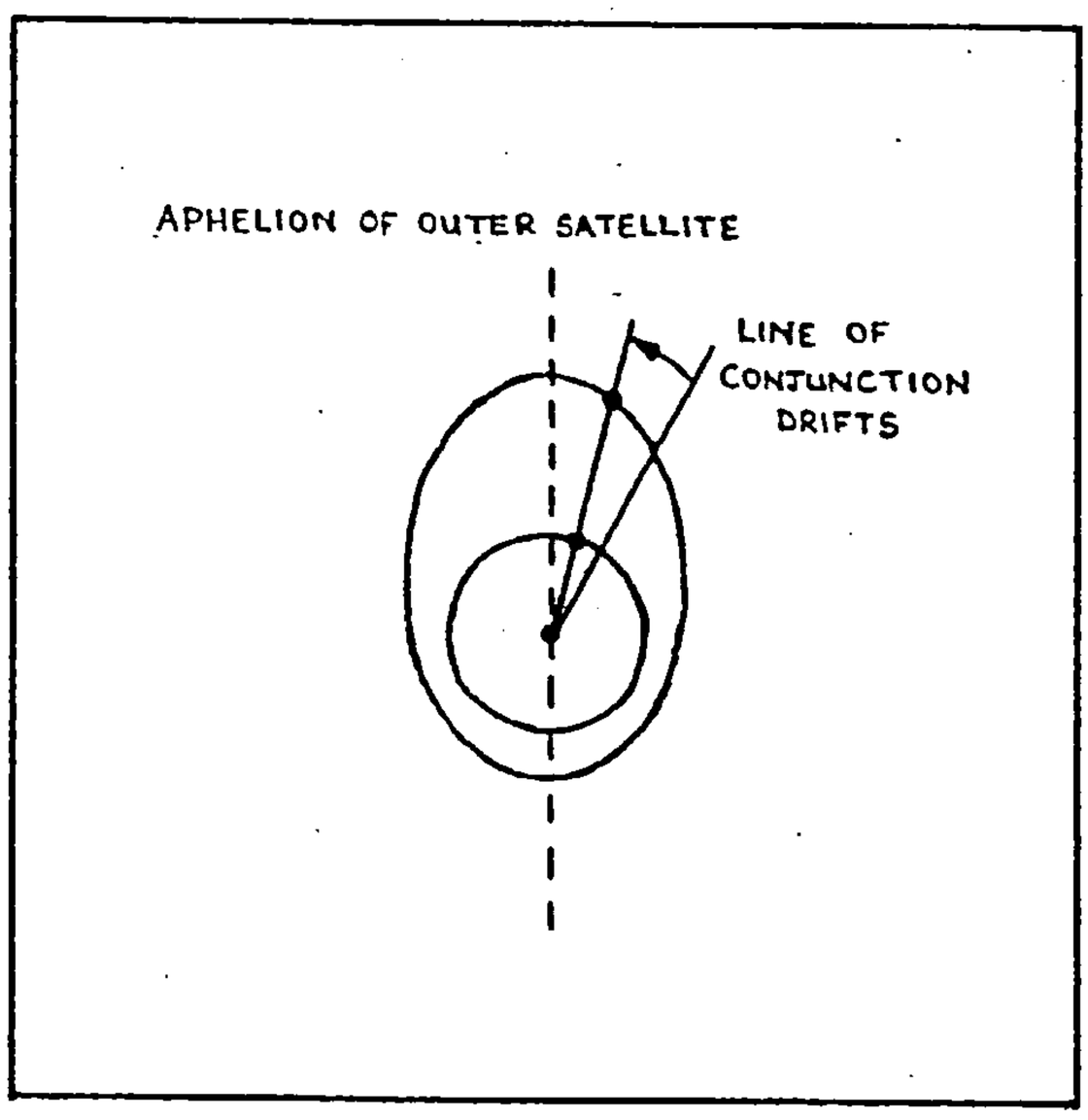


Figure 1.9(b): Transfer of angular momentum from the outer to the inner body causes the period of the outer to shorten and the inner to lengthen. Therefore, the line of conjunction drifts towards the apse line. After crossing the apse line, the flow of angular momentum is reversed. The drift of the line of conjunction is slowed down, halted and reversed.

1.4

satellites Miranda, Ariel and Umbriel, (see Lazzaro, Ferraz-Mello and Veillet, 1984).

It seems certain that tidal effects are responsible for the evolution of many satellite orbits. It is also likely that they are responsible for stabilising some satellite systems in commensurabilities. Tidal effects are too small to have an effect on planetary orbits, within the Solar System as it is presently arranged. During planetary formation, such effects were greater and may account for some of the resonances observed today (eg. Jupiter-Saturn is in a 5:2 near-commensurability).

1.5 General Perturbations and KAM Theory

This section is concerned with the study of Laplace's definition of stability by the use of general perturbations and Kolmogorov-Arnol'd-Moser (KAM) theory. Laplace's definition of stability in a n-body dynamical system requires that no collisions take place and no bodies escape the system.

This simple, qualitative definition can be applied to the Sun and major planets of the Solar System as well as a host of satellite systems. In the case of the Solar System, the only planet that does not come under the scope of Laplace's definition is of course Pluto. All the other planetary orbits are well spaced and non-crossing (recall Figure 1.1). The question then arises: will the status quo be maintained or will some orbits evolve so as to cross others?

Laplace himself made one of the first attempts to answer this question, by solving the Lagrange planetary equations (mentioned in

1.5

Section 1.4). They describe the rate of change of a set of orbital elements when that orbit is perturbed, these perturbations being characterised by the *disturbing function*. (It should be noted that the Lagrange planetary equations hold for many different classes of perturbations. Only the form of the disturbing function will alter).

Laplace found that the first order solution contained no secular terms in the expressions which yield the changes in the semi-major axes. The existence of secular terms would imply that the semi-major axes change linearly with time. This result therefore implies that each planet is restricted to an annulus containing its present orbit. The size of the periodic variations in the semi-major axes govern the widths of the annuli. Laplace demonstrated that the planetary annuli do not cross and are of small width in comparison to the orbits' mutual separations. He concluded that to first order, collisions and escapes were impossible.

In later years the second order theory was studied by Poisson and the third order by many others. At present the state of the problem is that, if expanded in a particular way, the theory shows no secular terms in the changes in semi-major axis to any order (Message, 1978). It should be pointed out that these results require that only point-mass Newtonian gravitational forces act on the bodies.

Laplace's approach was generalised by Newcomb (1876) who showed that the n-body problem admitted purely periodic solutions whose only secular terms arose in the angular variables. This assumed that the central mass was large compared to the others and that the others moved in near circular, coplanar orbits about the central mass. A few years later, Poincaré showed that the resulting trigonometrical series were in

1.5

general divergent, thus ruling out this approach for studying the long-term stability of the Solar System.

The interest in perturbation theories was revived in the middle of this century, principally by Kolmogorov (1954), followed by Arnol'd (1963) and Moser (1973). In their work (now known as KAM theory), they gave conditions under which the approach used by Laplace, Lagrange and Poisson could give rise to convergent series. With certain mathematical constraints, convergent series arise if:

(i) the perturbations within the system are sufficiently small.

(ii) the ratios of the natural frequencies of the system are poorly approximated by rational numbers.

Although KAM theory can be usefully employed for studying satellite orbits, it is unfortunately not applicable to the study of the stability of the Solar System. The second condition is satisfied except in cases of high resonance (not found amongst the planets). The first condition is not satisfied because the masses of the planets are too large in comparison to that of the Sun. It should be pointed out however that the limits on perturbations are lower bounds only. It may be that the KAM theory is valid for a larger range of perturbations which encompass those found between the planets. Another drawback is the fact that there exist solutions to the differential equations which fail the second condition of irrationality and are unstable, but exist arbitrarily close to well behaved stable solutions. It may turn out therefore that the KAM theory, while attractive in principle, is of little use in deciding the stability of the Solar System.

1.6 Special Perturbation Methods

We saw in the previous section, that general perturbations produce an elegant method for solving the equations of motion of a dynamical system. Unfortunately these analytical methods are often unusable in practice, due to such problems as small divisors and the need to go to high orders to achieve accuracy.

The alternative is to use special perturbation methods to solve the equations of motion by a numerical algorithm. Given the positions and velocities of the bodies at a particular epoch, it is possible to compute the displacements a short time later due to the forces each body exerts. Using these new updated positions and velocities, the procedure may be repeated many times in a step-by-step manner until numerical errors accumulate so much as to render any positions hopelessly inaccurate.

In choosing a numerical procedure for solving the equations of motion it is necessary to weigh carefully all factors before deciding. These factors may be grouped as follows:

(a) Orbit Type. Is the orbit in question near circular, highly elliptical or hyperbolic? Could it evolve drastically during the period of study?

(b) Operational Requirements. For how long is the integration to last? Is one long computation better than many smaller ones with different initial conditions? What accuracy requirements need to be imposed?

(c) Equations of Motion. Are they formulated as first-order or second-order differential equations?

1.6

(d) Numerical Algorithm. Which is preferable: a single-step or a multi-step method; a Cowell-type method or an Encke-type method? (These terms are explained below).

(e) Computing facilities. How sophisticated are the computing facilities at your disposal? (Do you have your own desk-top vector processor, or a pencil, notepad and your fingers?) How fast and how accurate are they? How much memory is required? How expensive is it?

Most of these questions can only be answered in the light of a particular problem. However it is worth considering two factors in more detail: accuracy and the choice of algorithm.

There are many factors which affect the accuracy of the results at the end of a computation. Some examples of sources of error are:

(i) truncation of infinite series to a finite number of terms. The error is the remainder at each step and may be negligible if enough terms are retained. This assumes that the series converges!

(ii) round off error. A computer holds only a finite number of digits for each number. Any arithmetic operation is rounded to that number of significant figures. Round off error is always present, but may be reduced by increasing the number of significant figures.

(iii) imperfect convergence. While performing any iteration, the convergence after a finite number of steps is not complete but can be made negligibly small.

(iv) the physical model. Relevant physical effects may be omitted from the theory. The initial conditions and masses may be in error.

(v) theoretical instability. Two nearby orbits may be in a region of chaos (Section 1.3) and diverge exponentially.

1.6

All these effects combine to cause inaccuracies which accumulate with each step of the integration. It is never clear how the various sources of error interact with each other. Workers usually minimise the errors as much as possible. They need to be aware however that an increase in the accuracy of an integration procedure is usually accompanied by a decrease in speed.

Numerical algorithms may be divided into single-step and multi-step methods. Multi-step methods require that the positions and velocities of the bodies from several previous steps be used to calculate the next step. Single-step methodsrequire only the last known positions and velocities. In general multi-step methods are faster, more stable and simpler to implement. Unfortunately they need a special procedure for generating the first few steps before the main integrator can start. The single-step method is preferred if frequent changes in the step size are necessary, for example when dealing with highly eccentric orbits or near collisions.

The other main classification is between Cowell-type and Encke-type methods. Cowell-type methods generally refer to methods where no knowledge of the orbital behaviour is assumed. Encke-type methods measure the differences between a real orbit and a fixed reference orbit (usually taken to be a Keplerian ellipse). At some epoch, when the differences become too great, the reference orbit is changed to that given by the present osculating elements of the real orbit. This process is called *rectification*. The advantages of the Encke methods over the Cowell methods are greater accuracy and a larger step size if the differences remain small. If the differences grow fast so that frequent rectification is required, a Cowell method may be preferable.

1.6

Also the Cowell methods are easier to implement and of wider application.

One final method worth mentioning is that of numerically integrating the Lagrange planetary equations. This may be done on a step-by-step basis as for other integration methods. Alternatively, the osculating elements at the start are substituted into the equations which are integrated over an extended period. This provides the first order perturbations which may be substituted into the equations. The procedure is then repeated to obtain the second order perturbations, and so on. This is a strict perturbation method of wide applicability. A larger step size may be used than is possible with rectangular coordinates. Unfortunately the method is unstable when the eccentricity of the orbit tends to zero or unity, or when the inclination tends to zero or π. Various changes of variable can be used to eliminate these problems, but this method is still difficult to implement.

There are many different numerical integration methods (Isaacson and Keller, 1966) and every worker has his/her favourite, which he/she feels is most appropriate for a particular problem. Fox (1982) provides a useful critique of some of the more popular methods.

At this point, it is worth discussing the use of Liapunov characteristic exponents in determining orbital stability from the results of a numerical integration. We wish to investigate the stability of the orbit described by the points P_o, P_1, P_2,... in Figure 1.10. At time $t = 0$, a body is at position P_o in its orbit.

At time $t = \tau$, it will be at P_1; at $t = 2\tau$ it will be at P_2, etc. If the body is displaced from P_o by an amount d, it will travel to a point P_1' after time τ which will be displaced from P_1 by an amount d_1.

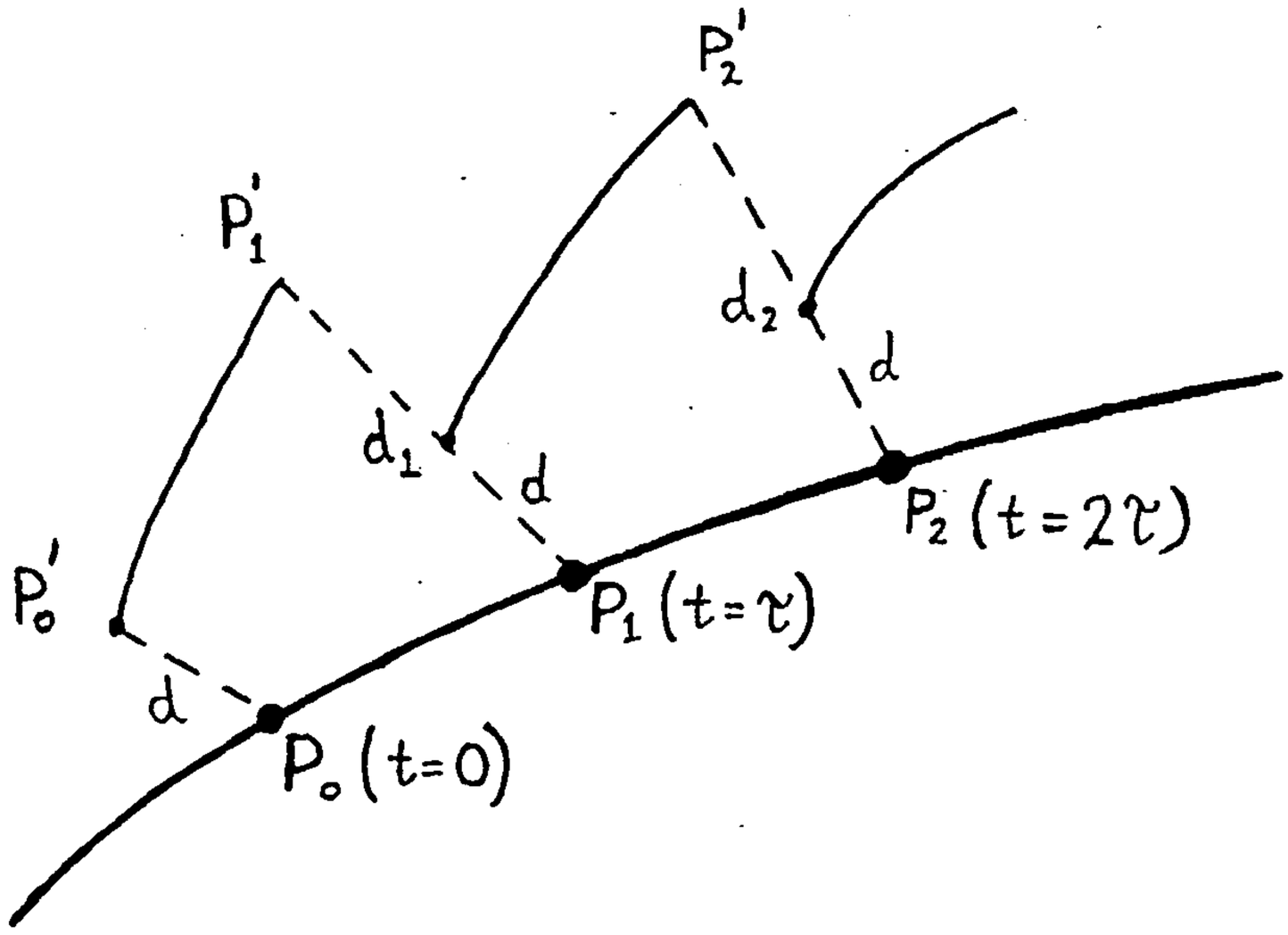


Figure 1.10: Numerical computation of a Liapunov characteristic exponent.

Similarly, if a body is displaced from P_1 by an amount d, the displacement at $t = 2\tau$, from P_2, will be d_2, and so on. The Liapunov characteristic exponent may be presented in the form

$$L = \lim_{n\to\infty} \frac{1}{n\tau} \sum_{i=1}^{n} \log \frac{d_i}{d}$$

If $L < 0$, the orbit is considered to be *asymptotically stable*.

If $L = 0$, the orbit is described as *quasi-periodic*. If $L > 0$, the orbit is *chaotic* (unstable). Although this stability criterion is widely applicable and gives useful information, it is sometimes difficult to decide the value of L in the limit.

1.6

We now consider some of the more noteable special perturbation studies concentrating on those investigating the stability of the Solar System. Three studies of the orbits of the five outer planets have been made by Cohen and Hubbard (1965), and Cohen, Hubbard and Oesterwinter (1967, 1972). The first integration spanned 120,000 years and showed the existence of a critical argument between Neptune and Pluto which prevents a close approach of less than 18 astronomical units (Section 1.4). The second study refined these results using improved elements for Pluto.

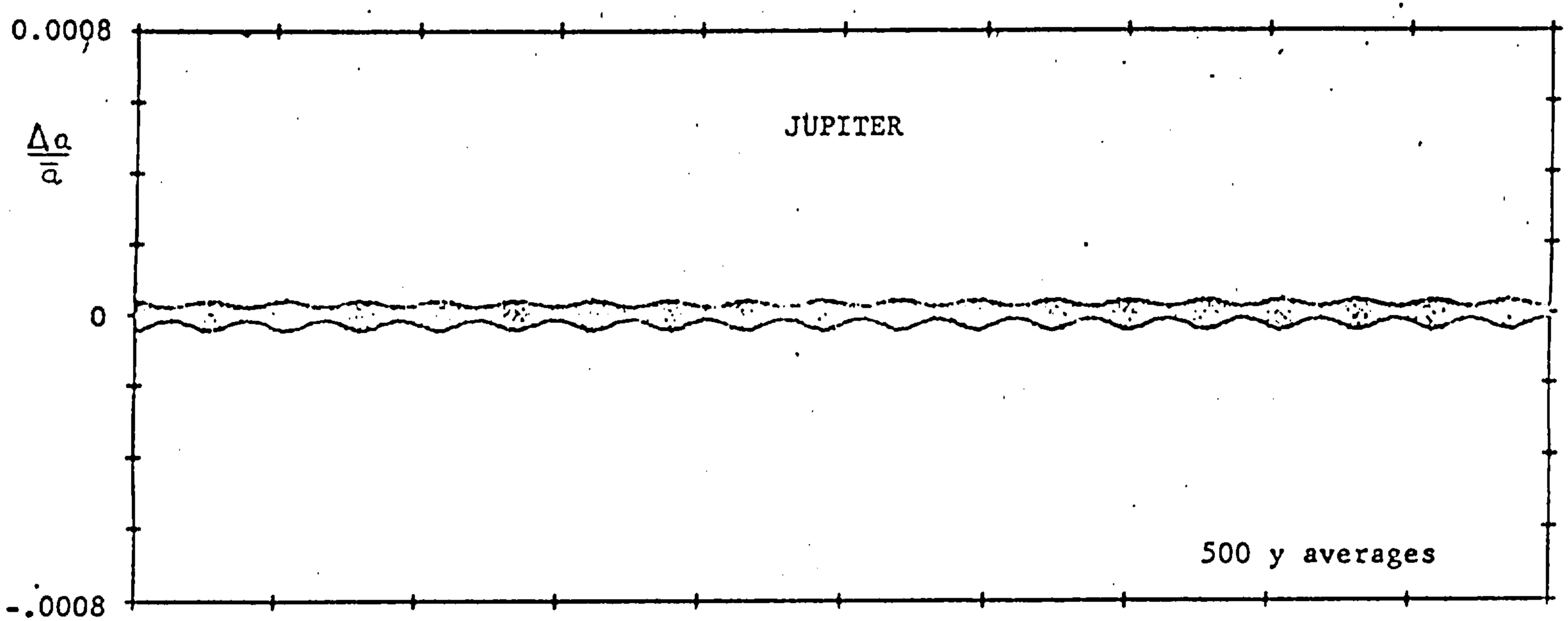
Their last study spanned 500,000 years forward and backwards from epoch Jan. 6.0 1941. They found no evidence of any secular trends in the elements a,e,i for any of the planets (Figure 1.11). The evidence was inconclusive for Pluto, which showed variations in the elements with a period of 19500 years. It was possible that the eccentricity and inclination possessed secular trends.

Williams and Benson (1971) performed a numerical integration of Pluto over a period of 4.5×10^6 years to look for secular or resonant behaviour. Assuming that the other four (outer) planets' orbits were known, they eliminated the short period terms in Pluto's orbit, using the method of Gauss secular variations. They found that the argument of perihelion librates about 90^o with an amplitude of 24^o in a period of 3.995×10^6 years.

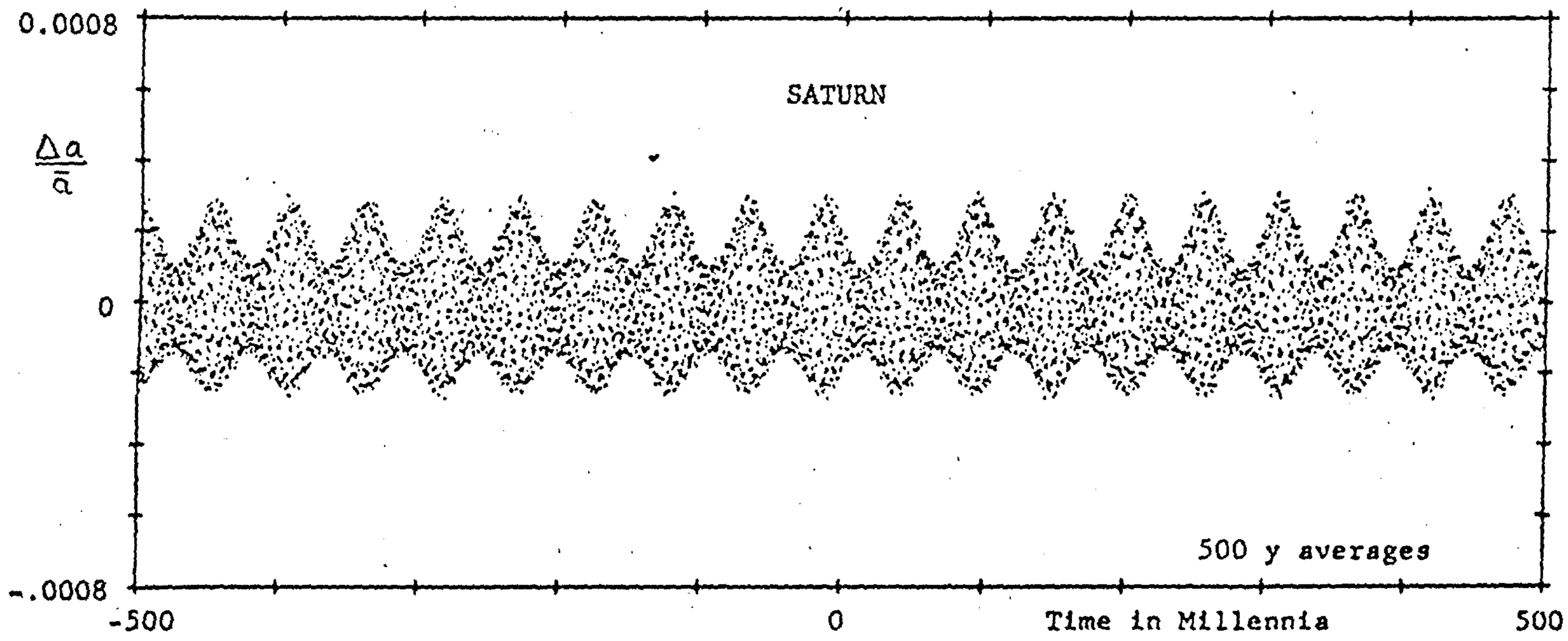
More recently , Kinoshita and Nakai (1984) integrated the equations of motion for the outer planets over a period of 5×10^6 years. These results have been analysed by Milani and Nobili (1984) who showed that the angle between the perihelia of Jupiter and Saturn librate around 180^o with an amplitude of 70^o and a period of 1.1×10^6 years.

Figure 1.11: Results from Cohen et al. (1972) (a) Relative variation in semi-major axis for Jupiter, $\bar{a}$ = 5.20257 A.U.
(b) Relative variation in semi-major axis for Saturn, $\bar{a}$ = 9.5549 A.U.
(c) Variation in eccentricity for Jupiter and Saturn

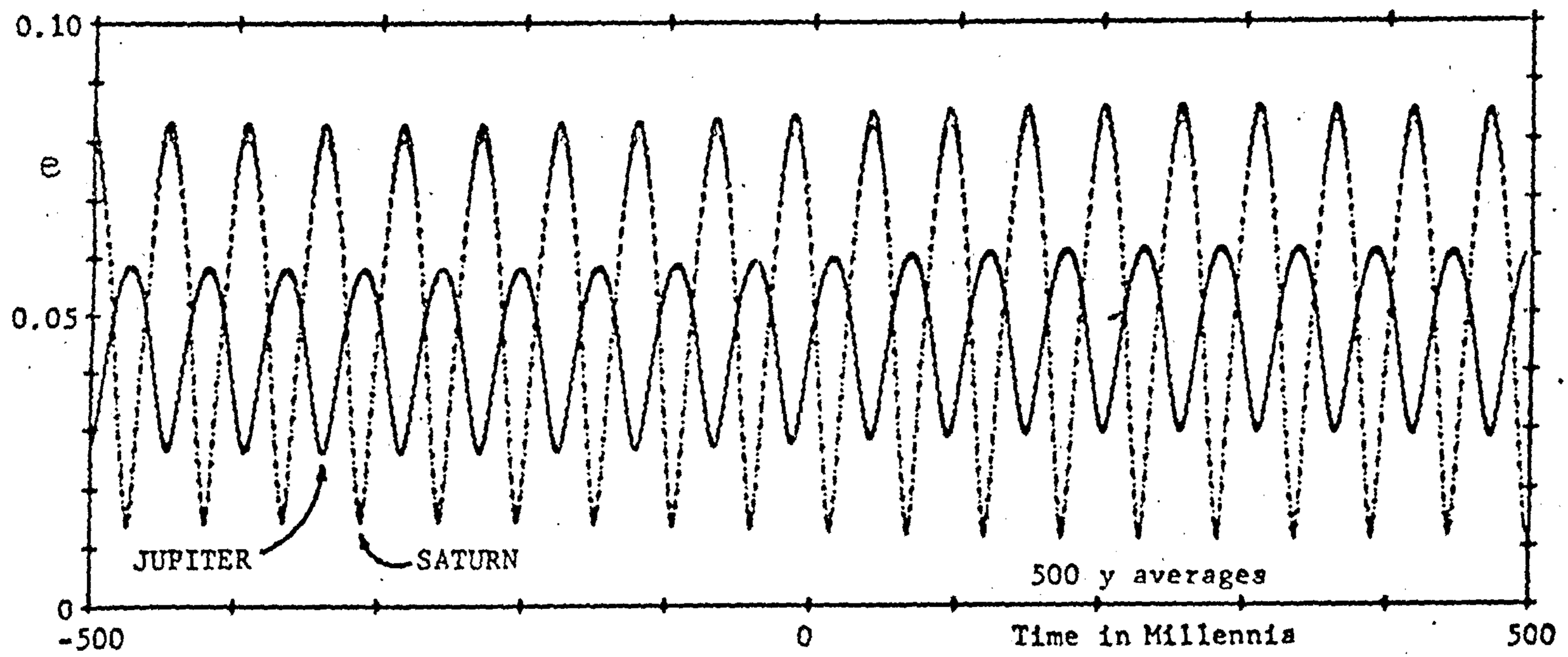
(a)



(b)


0.0008
SATURN
Δa/ā
0
-.0008
500 y averages
-500
0
Time in Millennia
500

(c)


0.10
e
0.05
0
JUPITER
SATURN
500 y averages
-500
0
Time in Millennia
500

1.6

Most recently, a project is in progress to study the behaviour of the outer planets over a period of 10^8 years, (Roy, 1983). Studies involving all nine planets span a much smaller time, the longest being 4400 years (JPL DE102). This is due primarily to the increase in the number of bodies making the integration time per step longer. It is also due to the greater speed of the inner bodies requiring a small step size with the possible inclusion of relativistic effects.

Finally we consider some studies of fictitious systems. Nacozy (1977) has studied the stability of the Sun-Jupiter-Saturn system by integrating the equations of motion with augmented Jupiter and Saturn masses. He found that if the masses of the planets exceeded their real values by a factor of around thirty, then secular trends appeared in the semi-major axes and eccentricities. He implied that the real masses, being much smaller, should give rise to stable orbits.

Harrington (1972,1977) has considered several classes of hierarchical three-body systems (one body widely separated from the other two, as described in Chapter 2). Consider Figure 1.12 showing three bodies S_1, S_2, S_3 with masses m_1, m_2, m_3 respectively. Harrington studied:

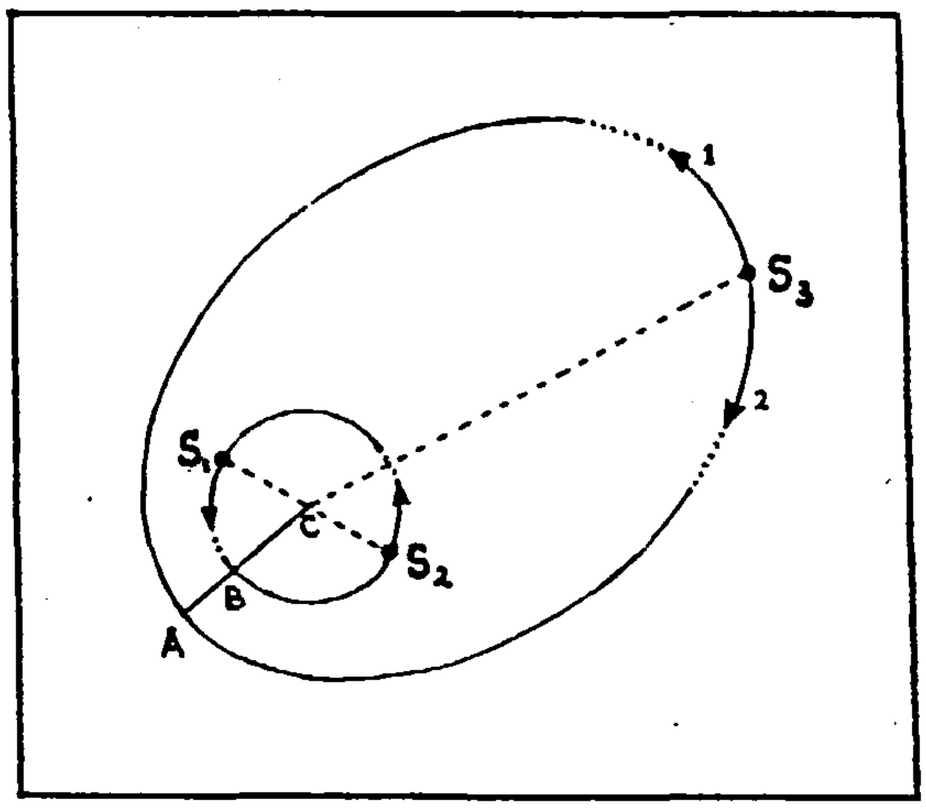
(a) triple stellar systems, ($m_1 \sim m_2 \sim m_3$).

(b) inferior planets in double star systems ($m_2 << m_1 \sim m_3$)

(c) superior planets in double star systems ($m_3 << m_1 \sim m_2$).

For class (a), Harrington found that the stability varied depending on the rotation of the third body with respect to the other two, i.e. direct (arrow 1 in Figure 1.12) or retrograde (arrow 2). The stability also varied with the ratio of the pericentre distance of the outer orbit to the semi-major axis of the inner orbit (AC:BC in Figure 1.12).

1.6



S_1, S_2, S_3 denote stars.
C is the mass-centre of the close binary.
A is the position of the pericentre of S_3.

Figure 1.12:

In the case of equal mass triple stellar systems, it was found that this ratio must exceed 3.5 for direct systems and 2.75 for retrograde systems to ensure that no irreparable changes occur to the overall structure of the systems.

For inferior and superior planets (b), (c) in double star systems, stability seemed assured if the ratio AC:BC exceeded a value between 3 and 4, regardless of rotational sense. Harrington's results are extremely relevant to the following work and are discussed again in Chapter 6.

A much more thorough study of direct hierarchical three-body systems has been performed by Walker and Roy, using empirical stability parameters. Their formulation of the three-body problem as well as many of their results form the basis for the work presented in subsequent chapters. Their work is therefore reviewed in more detail within Chapters 2,4 and 5.

1.7 Summary

The purpose of this chapter has been to introduce various definitions

1.7

of stability and comment on their usefulness. Many of these definitions are relevant to the work presented in subsequent chapters. Examples of their applications have mostly been taken from the Solar System. Many questions were asked in Section 1.1 concerning the stability of bodies in the Solar System. Few if any can be answered with much confidence given the studies to date. The main problem is that the conclusions drawn from a particular study are only as valid as the physical model and stability definition used. For example, if the model uses point-mass bodies and Newtonian gravity, critical effects due to tidal interactions (Section 1.4) and general relativity may be missed, (there is a discrepancy between classical theory and observation in the precession rate of Mercury's apse which may be remedied by a relativistic correction). These effects may not be negligible over the age of the Solar System ($\sim 5 \times 10^9$ years).

Given that it is impossible to say conclusively whether a real system is stable or not, another question may be asked: how likely is it, that the system will become unstable during a given length of time? Unless we are content to use phrases such as "extremely", "possible", or "not very", the help of probability and statistical theory should be sought, to give a quantitative estimate. With the advent of high speed computers, it is now possible to generate large amounts of data by numerical integration. For the first time, the worker has the opportunity to gather enough results in his lifetime to make a thorough statistical examination.

In subsequent chapters, it is shown how simple statistical tools can be applied to the results of numerical integration experiments. Quantitative predictions of stability lifetimes in three-body systems

1.7

are obtained, and a method for observing resonances is described.

The work presented here concentrates on the stability of hierarchical systems. The definitions of a *hierarchical system* and *hierarchical stability* as given by Walker and Roy are presented in Chapter 2. The definition of hierarchical stability is Laplace-like in its qualitative nature; nevertheless quantitative criteria can be derived from it.

Hill-type stability (Section 1.2) can be applied to the general three-body problem (Zare, 1976, 1977; Marchal and Saari, 1975) and has been used in the study of real and fictitious systems (Szebehely, 1977; Szebehely and McKenzie, 1977 a,b; Szebehely and Zare, 1977). This definition, reviewed in Chapter 3, forms one of the conditions for the more stringent definition of hierarchical stability.

In Chapters 4 and 5, special perturbation methods are used to study general three-body systems. All the systems are fictitious, coplanar and maybe direct or retrograde. Statistical techniques are used to analyse these results. Particular attention is paid to the influence of commensurabilities (Section 1.4).

In Chapter 6, the results for the fictitious systems are compared with real systems and with the results of Harrington (Section 1.6). Harrington showed that retrograde orbits were generally more stable than the corresponding direct orbits. The range of parameters that he investigated was rather small and he performed fewer experiments than are carried out here. The results of this work confirm his findings and mirror the results of Hénon who found similar results for the restricted three-body problem, (Section 1.3). For the restricted *and* general problems, the numerical results contradict the results of a Hill stability analysis, which shows that retrograde systems are less stable than direct systems.

1.7

Finally, in Chapter 7, commensurabilities and mirror configurations for four and more body systems are examined. In planetary systems, a statistical analysis of the numbers of syzygies (alignments) of the bodies in a given time, leads to the discovery of resonances. The theory for this method is developed and its application to the results of numerical experiments is discussed.

CHAPTER 2

EMPIRICAL STABILITY PARAMETERS IN HIERARCHICAL DYNAMICAL SYSTEMS

2.1 Introduction

Celestial mechanics is concerned with determining the relative motion of n bodies that are gravitationally interacting in three-dimensional space. In general, the problems encountered are insoluble by purely analytical methods. It is usually necessary to make certain simplifying assumptions. The basic assumptions expressed in this chapter will be used throughout all the chapters in this thesis.

The first simplification that is made is to treat all bodies according to Newton's Law of Gravity and neglect any relativistic effects. The accuracy of this assumption will be governed by the speed of the bodies in question as well as by the duration of study of the system.

The second simplification is that only point masses are considered. This means that we will neglect any effects due to the finite size, irregularity of shape or non-uniformity in internal distribution of mass of the bodies. If the bodies have spherically symmetric mass distributions then they may be exactly represented by point masses. This is a reasonable assumption for most bodies in the Solar System with the exception of some satellites and asteroids.

Thirdly, we will be concerned only with two-dimensional motion. This assumption of coplanarity is reasonable for the Solar System and most satellite systems. The most notable exceptions to this rule, are Pluto, (inclined at 17^{o} to the ecliptic) and some of the Jovian satellites.

The fourth assumption is that the system is hierarchical in structure. A full description of this concept is given in Section 2.2 as well as a definition of hierarchical stability that will be used throughout this work.

2.1

In Section 2.3, Jacobian coordinates are discussed. The derivation of the equations of motion for n bodies in Jacobian coordinates is given in Section 2.4. The particular case of three bodies is discussed in more detail in Section 2.5.

The contents of this thesis rely heavily on definitions and results taken from the work of I. W. Walker and A. E. Roy. In this chapter, some of the relevant concepts are explained. More detailed explanations are given in Walker (1980) and a summary is available in Roy (1979). Much of their work is explained in five papers by them, written between 1980 and 1983. In future these will be referred to as Walker and Roy I, II, III, IV and V.

2.2 Hierarchical Systems and their Stability

A dynamical n-body system is described as hierarchical if, when described by a suitable coordinate system, the orbital radii may be ordered in ascending size and that order is maintained for a time interval at least as long as the longest orbital period in the system. As we shall see in the next section, a Jacobian coordinate system is preferred in dealing with hierarchical systems (Roy et al. (1985)).

Evans (1968) described the hierarchical arrangement of systems by "mobile diagrams", (Figures 2.1 (a)-(c)). Figure 2.1(a) shows a planetary system where each planet is successively further from the sun. Satellite systems around a planet can be characterised in the same way, and Figure 2.1(b) shows the arrangement when both planetary and satellite systems are given in the same diagram. Multiple star systems can take up very complex arrangements, depending on the number

2.2

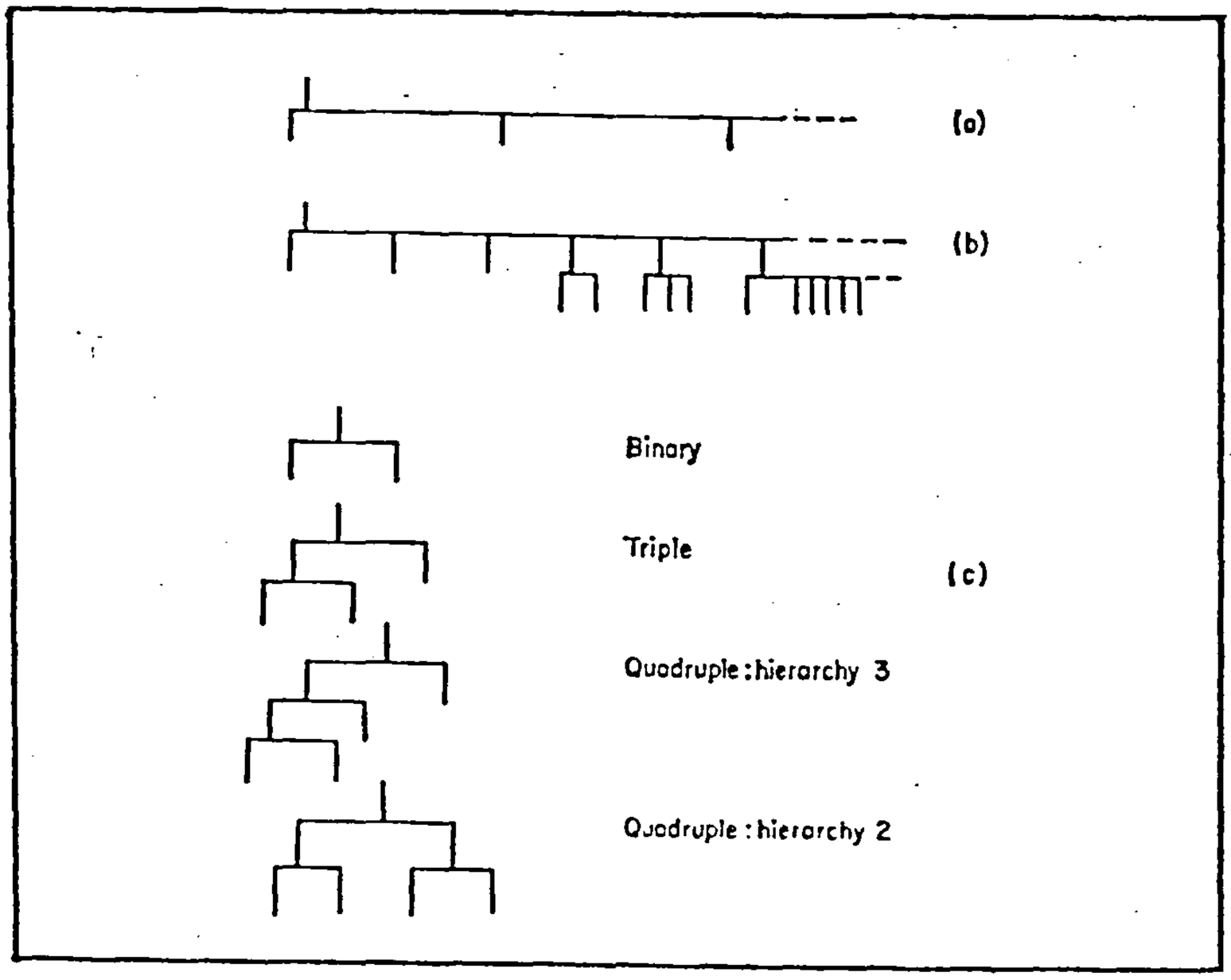


Figure 2.1 Evans mobile diagrams for (a) planets (b) planets and satellites and (c) multiple stellar systems.

of bodies (Figure 2.1(c)). Binary and triple systems can be arranged in only one way. A quadruple system may be arranged in two distinct ways. The simple hierarchy (3) is equivalent to the planetary hierarchy while the double binary hierarchy (2) has two binary systems orbiting a common centre of mass.

Complex hierarchical arrangements may be broken down into their constituent simple hierarchies (i.e. hierarchies where the bodies are successively further away from the first (principal) body). In this case a "body" may refer to the centre of mass of a subsystem. A nested set of numbers can describe any hierarchy, by breaking it up into its constituent simple hierarchies, (Walker and Roy V). For example, the simple four-body system would be a 4-system, while the double binary would be a 4(2,2)-system. The Solar System of Sun, planets and major satellites (54 bodies in all) is described as

2.2

54(52(49(43(25(8(5(3,2),3),17),18),6),3),2).

The (3,2) system is the triplet of Sun-Mercury-Venus with the Earth-Moon binary. This 5-body system forms a binary with the Martian system (3 bodies). This 8-body system forms a binary with the Jovian system, and so on.

In all future work, we shall restrict ourselves to examining only simple hierarchies and in particular, three-body hierarchical systems, where two of the bodies form a close binary relative to the third body.

Walker and Roy III give the following definition of hierarchical stability. A hierarchical dynamical n-body system is held to be stable if, during an interval of time substantially longer than the periods of revolution of the bodies in the system:

(i) none of the bodies escapes from the system;

(ii) the ordering of the sizes of the radius vectors that define the hierarchy remains unchanges;

(iii) no irreversible changes occur in any orbit's size, shape or orientation to the invariable plane of the system.

These conditions ensure that the perturbed binary orbits that make up the system, undergo no drastic changes.

This is the definition of stability that will be used throughout the rest of this work and will be explained in more detail in Chapter 4 when the numerical experiments will be discussed.

2.3 Jacobian Coordinates

The Jacobian coordinate system and its application to the general

2.3

n-body problem was introduced by Jacobi and Lagrange (Plummer, 1918). This coordinate system is particularly useful for hierarchical dynamical systems where we are interested in the separation of a body from the centre of mass of an adjacent subsystem.

Consider an n-body system with masses $m_1, m_2, \ldots, m_n$, (see Figure 2.2). With respect to an inertial reference frame (origin O) the equations of motion are

$$m_i \ddot{\underset{\sim}{R}}_i = \nabla_{\underset{\sim}{i}} U \qquad (i=1,2,\ldots,n) \qquad (1)$$

where $$U = \frac{1}{2} G \sum_{k=1}^{n} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} \frac{m_k m_\ell}{r_{k\ell}} \qquad (2)$$

is the force function.

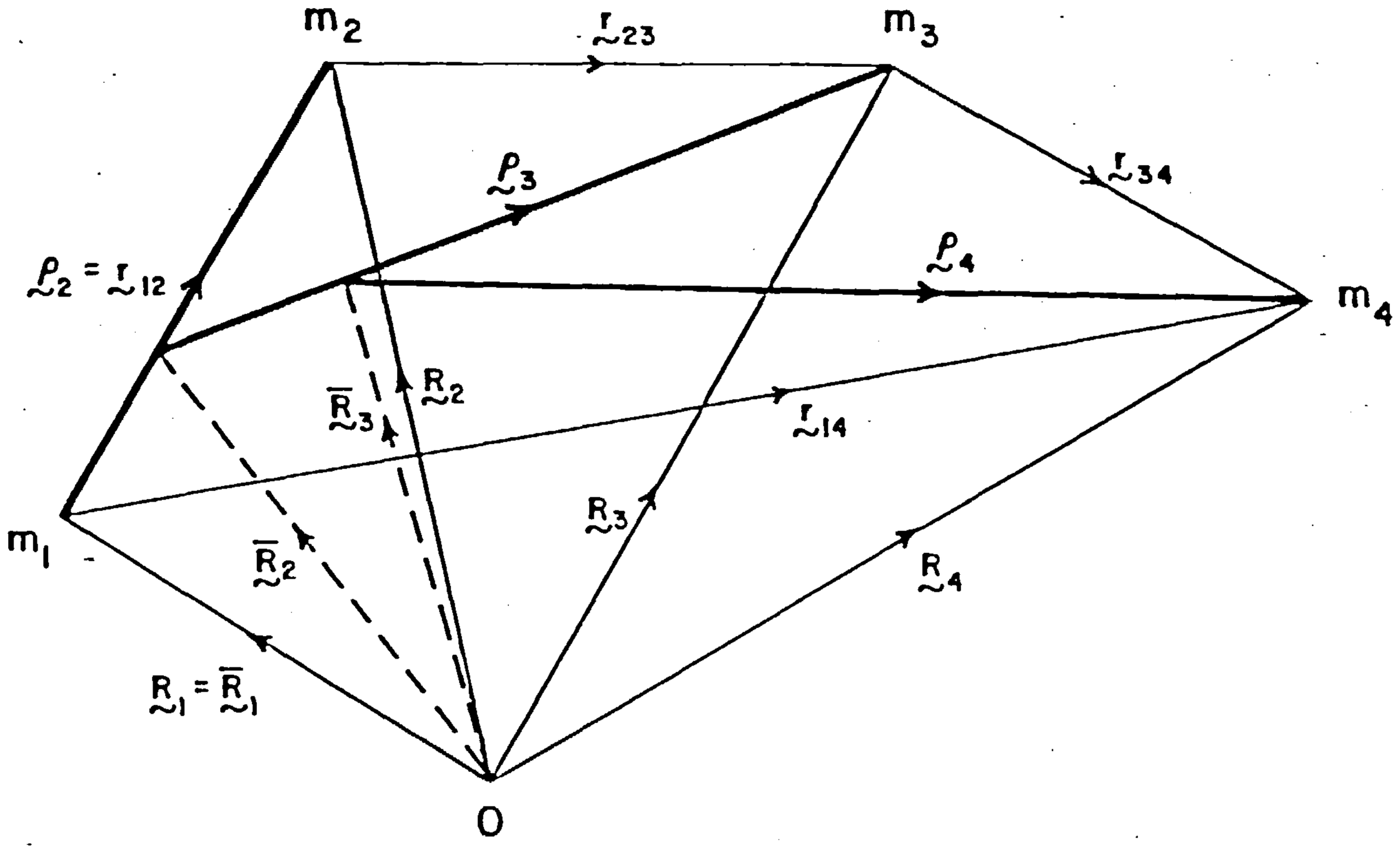


Figure 2.2 A particular case of the Jacobian coordinate system where n=4. (see text for definition of symbols used).

2.3

$\underset{\sim}{R}_i$ is the position vector $\overrightarrow{Om}_i$

$\underset{\sim}{\nabla}_i$ is the gradient operator with respect to $\underset{\sim}{R}_i$

$r_{k\ell} = |\underset{\sim}{r}_{k\ell}| = |\underset{\sim}{R}_\ell - \underset{\sim}{R}_k|$

G is the gravitational constant,

and a dot denotes differentiation with respect to time.

Let us suppose that the bodies lie in a simple hierarchical arrangement. This means that successive bodies m_i (i=2,...,n) lie at a greater distance than the previous (i-1) bodies, from the centre of mass of the (i-1)-body subsystem. The vector connecting this centre of mass to m_i is denoted by $\underset{\sim}{\rho}_i$ and is called the Jacobian radius vector. The $\underset{\sim}{\rho}$'s form the basis of the Jacobian coordinate system. It is now necessary to express the equations of motion in terms of the new coordinates $\underset{\sim}{\rho}_i$.

The centre of mass of the first j masses has position vector $\underset{\sim}{\bar{R}}_j$, where

$$M_j \underset{\sim}{\bar{R}}_j = \sum_{k=1}^{j} m_k \underset{\sim}{R}_k \qquad (j=1,\ldots,n) \qquad (3)$$

and

$$M_j = \sum_{k=1}^{j} m_k \qquad (j=1,\ldots,n) \qquad (4)$$

From the definition,

$$\underset{\sim}{\rho}_i = \underset{\sim}{R}_i - \underset{\sim}{\bar{R}}_{i-1} \qquad (i=2,\ldots,n)$$

i.e.

$$\underset{\sim}{\rho}_i = \underset{\sim}{R}_i - \frac{1}{M_{i-1}} \sum_{k=1}^{i-1} m_k \underset{\sim}{R}_k \qquad (i=2,\ldots,n) \qquad (5)$$

We differentiate Equation (5) twice, multiply it by m_i, and substitute Equation (1) in, to get

$$m_i \ddot{\underset{\sim}{\rho}}_i = \nabla_{\underset{\sim}{i}} U - \frac{m_i}{M_{i-1}} \sum_{k=1}^{i-1} \nabla_{\underset{\sim}{k}} U \qquad (i=2,\ldots,n) \qquad (6)$$

Let $\underset{\sim}{R}_i = (X_i, Y_i, Z_i)$, $\underset{\sim}{\rho}_i = (\xi_i, \eta_i, \zeta_i)$, in the inertial frame.

The x-component of Equation (6) is therefore

$$m_i \ddot{\xi}_i = \frac{\partial U}{\partial X_i} - \frac{m_i}{M_{i-1}} \sum_{k=1}^{i-1} \frac{\partial U}{\partial X_k} \qquad (i=2,\ldots,n) \qquad (7)$$

From the x-component of Equation (5), we see that

$$\frac{\partial \xi_j}{\partial X_i} = \begin{cases} 0 & \text{if} \quad j < 1 \qquad (j=2,\ldots,n) \\ 1 & \text{if} \quad j = 1 \qquad (i=1,\ldots,n) \\ -\dfrac{m_i}{M_{j-1}} & \text{if} \quad j > i \end{cases}$$

Hence

$$\frac{\partial U}{\partial X_i} = \sum_{j=2}^{n} \frac{\partial U}{\partial \xi_j} \frac{\partial \xi_j}{\partial X_i} \qquad (i=1,\ldots,n)$$

$$= \frac{\partial U}{\partial \xi_i} - m_i \sum_{j=i+1}^{n} \frac{1}{M_{j-1}} \frac{\partial U}{\partial \xi_j} \qquad (i=1,\ldots,n)$$

where we define $\partial U/\partial \xi_1 = 0$. Equation (7) becomes

$$m_i \ddot{\xi}_i = \frac{\partial U}{\partial \xi_i} - m_i \sum_{j=i+1}^{n} \frac{1}{M_{j-1}} \frac{\partial U}{\partial \xi_j} - \frac{m_i}{M_{i-1}} \sum_{k=1}^{i-1} \frac{\partial U}{\partial \xi_k}$$

$$+ \frac{m_i}{M_{i-1}} \sum_{k=1}^{i-1} \sum_{j=k+1}^{n} \frac{m_k}{M_{j-1}} \frac{\partial U}{\partial \xi_j} \qquad (8)$$

By changing the order, the double summation becomes

2.3

$$\frac{m_i}{M_{i-1}} \sum_{k=1}^{i-1} \sum_{j=k+1}^{n} \frac{m_k}{M_{j-1}} \frac{\partial U}{\partial \xi_j} = \frac{m_i}{M_{i-1}} \sum_{j=2}^{i} \frac{1}{M_{j-1}} \frac{\partial U}{\partial \xi_j} \sum_{k=1}^{j-1} m_k$$

$$+ \frac{m_i}{M_{i-1}} \sum_{j=i+1}^{n} \frac{1}{M_{j-1}} \frac{\partial U}{\partial \xi_j} \sum_{k=1}^{i-1} m_k$$

$$= \frac{m_i}{M_{i-1}} \sum_{j=2}^{i} \frac{\partial U}{\partial \xi_j} + m_i \sum_{j=i+1}^{n} \frac{1}{M_{j-1}} \frac{\partial U}{\partial \xi_j}$$

using Equation (4). Hence Equation (8) becomes

$$m_i \ddot{\xi}_i = \frac{\partial U}{\partial \xi_i} + \frac{m_i}{M_{i-1}} \left[\sum_{j=2}^{i} \frac{\partial U}{\partial \xi_j} - \sum_{k=1}^{i-1} \frac{\partial U}{\partial \xi_k} \right]$$

Since $\partial U/\partial \xi_1 = 0$,

$$m_i \ddot{\xi}_i = \frac{\partial U}{\partial \xi_i} + \frac{m_i}{M_{i-1}} \frac{\partial U}{\partial \xi_i} = \frac{M_i}{M_{i-1}} \frac{\partial U}{\partial \xi_i}$$

With similar results for the y and z-components, we have the equations of motion for n bodies in Jacobian coordinates,

$$\frac{m_i M_{i-1}}{M_i} \ddot{\underset{\sim}{\rho}}_i = \nabla_{\underset{\sim}{i}} U \qquad (i=2,\ldots,n) \qquad (9)$$

where $\nabla_{\underset{\sim}{i}}$ now denotes the gradient operator with respect to $\underset{\sim}{\rho}_i$.

Equations (9) form a $(6n-6)^{th}$ order system. This is a reduction from the original $6n^{th}$ order system, since we have exchanged the inertial frame for one moving with the centre of mass of the system.

In the next section we derive the force function U in terms of ratios ρ_i/ρ_j.

2.4 Expansion of the Force Function of a Hierarchical n-Body Dynamical System.

This section is a summary of work performed by Walker who describes it in much greater detail, (Walker, 1980). The notation of Section 2.3 still applies.

From Equation (2), U may be rewritten as

$$U = G \sum_{\ell=2}^{n} m_\ell \left[\sum_{k=1}^{\ell-1} \frac{m_k}{r_{k\ell}} \right] \qquad (10)$$

Thus it is required to find $1/r_{k\ell}$ in terms of the Jacobian radii.

Using equation (3) twice with subscripts j and j-1, we get

$$\bar{R}_j - \bar{R}_{j-1} = \frac{m_j}{M_j} (R_j - \bar{R}_{j-1}) \qquad (j=2,\ldots,n)$$

Summing this over j from k to ℓ-1 ($k < \ell$) yields

$$\bar{R}_{\ell-1} - \bar{R}_{k-1} = \sum_{j=k}^{\ell-1} \frac{m_j}{M_j} (R_j - \bar{R}_{j-1}) \qquad (11)$$

Now

$$r_{k\ell} = R_\ell - R_k = (R_\ell - \bar{R}_{\ell-1}) - (R_k - \bar{R}_{k-1}) + (\bar{R}_{\ell-1} - \bar{R}_{k-1})$$

Thus, using equation (11) and defining $\rho_1 = 0$,

$$r_{k\ell} = \rho_\ell - \rho_k + \sum_{j=k}^{\ell-1} \frac{m_j}{M_j} \rho_j . \qquad (1 \leq k < \ell \leq n)$$

On constructing $(r_{k\ell} \cdot \bar{r}_{k\ell})^{-\frac{1}{2}}$, we obtain

$$\frac{1}{r_{k\ell}} = \frac{1}{\rho_\ell} \cdot \left[1 + \alpha_{k\ell}^2 + \sum_{j=k}^{\ell-1} \sum_{h=k}^{\ell-1} \left[\frac{m_j m_h}{M_j M_h} \alpha_{j\ell} \alpha_{h\ell} C_{jh} \right] - 2\alpha_{k\ell} C_{k\ell} + 2 \sum_{j=k}^{\ell-1} \frac{m_j}{M_j} \alpha_{j\ell} (C_{j\ell} - \alpha_{k\ell} C_{jk}) \right]^{-\frac{1}{2}} \qquad (12)$$

2.4

where

$$\alpha_{ij} = \frac{\rho_i}{\rho_j}$$

$$C_{ij} = \frac{\underset{\sim}{\rho}_i \cdot \underset{\sim}{\rho}_j}{\rho_i \ \rho_j} \qquad (i < j) \qquad (13)$$

$$\rho_i = \left| \underset{\sim}{\rho}_i \right|$$

Note that C_{ij} is just the cosine of the angle between the i^{th} and j^{th} Jacobian radius vector. Note also that up to this point, there has been no assumptions made concerning any hierarchical structure. Now, however, it is desirable to apply a binomial expansion of the form $(1 + x)^{-\frac{1}{2}}$ to Equation (12) and truncate the expression after the second order in the α's. In order for this to be a valid procedure, we must now impose a hierarchical structure on the system. The conditions are

$$\left.\begin{array}{ll} \left[\dfrac{r_{k\ell}}{\rho_\ell}\right]^2 & -1 < 1 \\ \\ \alpha_{k\ell} & < 1 \end{array}\right\} \quad \forall k, \ell \qquad (k<\ell)$$

These conditions will be satisfied for a simple hierarchy of the form given in Figure 2.2, with $m_1 \geq m_2$, and the Jacobian radius vector increasing in size as we move out from m_1 towards m_n. With these assumptions,

$$\frac{1}{r_{k\ell}} = \frac{1}{\rho_\ell}\left[1 - \frac{1}{2}\alpha_{k\ell}^2 - \frac{1}{2}\sum_{j=k}^{\ell-1}\sum_{h=k}^{\ell-1}\left[\frac{m_j}{M_j}\frac{m_h}{M_h}\ \alpha_{j\ell}\ \alpha_{h\ell}\ C_{jh}\right] + \right.$$

$$+ \alpha_{k\ell}\ C_{k\ell} + \sum_{j=k}^{\ell-1}\frac{m_j}{M_j}\alpha_{j\ell}\ (\alpha_{k\ell}\ C_{jk} - C_{j\ell}) + \frac{3}{2}\alpha_{k\ell}^2\ C_{k\ell}^2 +$$

$$\left. + \frac{3}{2}\sum_{j=k}^{\ell-1}\sum_{h=k}^{\ell-1}\left[\frac{m_j}{M_j}\frac{m_h}{M_h}\ \alpha_{j\ell}\ \alpha_{h\ell}\ C_{j\ell}\ C_{h\ell}\right] - 3\alpha_{k\ell}\ C_{k\ell}\sum_{j=k}^{\ell-1}\frac{m_j}{M_j}\alpha_{j\ell}\ C_{j\ell}\right]$$

$$+ O(\alpha_{k\ell}^3) \qquad (14)$$

2.4

<u>Lemma 1</u>
$$\sum_{k=1}^{\ell-1} \frac{m_k}{\rho_\ell}\left[\alpha_{k\ell}\, C_{k\ell} - \sum_{j=k}^{\ell-1} \frac{m_j}{M_j}\, \alpha_{j\ell}\, C_{j\ell}\right] = 0 \qquad \forall \ell = 2,\ldots,n.$$

<u>Lemma 2</u>
$$\sum_{k=1}^{\ell-1} \frac{m_k}{\rho_\ell}\left[\frac{1}{2}\sum_{j=k}^{\ell-1}\sum_{h=k}^{\ell-1} \frac{m_j}{M_j}\frac{m_h}{M_h}\,\alpha_{j\ell}\,\alpha_{h\ell}\,(3C_{j\ell}\,C_{h\ell} - C_{jh}) - \right.$$
$$\left. - \sum_{j=k}^{\ell-1} \frac{m_j}{M_j}\,\alpha_{j\ell}\,\alpha_{k\ell}\,(3C_{j\ell}\,C_{k\ell} - C_{jk})\right] = -\frac{1}{\rho_\ell}\sum_{j=1}^{\ell-1} \frac{m_j^2}{M_j}\,\alpha_{j\ell}\,P_2(C_{j\ell})$$

$$\forall \ell = 2,\ldots,n.$$

where

$$P_2(x) = \frac{1}{2}\,(3x^2-1)$$

is the Legendre polynomial of order two.

<u>Proofs</u> See Walker (1980).

We substitute Equation (14) into Equation (10) and make use of Lemmas 1 and 2, to obtain, (after a little reduction)

$$U = G \sum_{\ell=2}^{n} \frac{m_\ell}{\rho_\ell}\left[M_{\ell-1} + \sum_{k=1}^{\ell-1} \frac{m_k\, M_{k-1}}{M_k}\,\alpha_{k\ell}^2\, P_2(C_{k\ell}) \right] + O\,(\alpha_{k\ell}^3) \,. \tag{15}$$

We use this expanded form of U to obtain the equations of motion in Jacobian coordinates. Thus, by Equation (9), noting the independence of the $\underset{\sim}{\rho}$'s,

$$\ddot{\underset{\sim}{\rho}}_i = G\, M_i\, \nabla_i \left[\frac{1}{\rho_i}\left(1 + \sum_{k=1}^{i-1} \varepsilon^{ki}\, P_2(C_{ki}) + \sum_{\ell=i+1}^{n} \varepsilon_{\ell i}\, P_2\,(C_{i\ell})\right)\right] \tag{16}$$

$$(i = 2,..,n).$$

2.4

where

$$\left.\begin{aligned} \varepsilon^{ki} &= \frac{m_k \; M_{k-1}}{M_{i-1} \; M_k} \alpha_{ki}^2 \qquad (1 \leqslant k < i \leqslant n) \\ \varepsilon_{\ell i} &= \frac{m_\ell}{M_i} \alpha_{i\ell}^3 \qquad (1 \leqslant i < \ell \leqslant n) \end{aligned}\right\} \tag{17}$$

where we define $M_o = 0$ and $\underset{\sim}{\rho}_1 = \underset{\sim}{0}$.

On examining Equation (16), it is seen that the first term on the right hand side represents the undisturbed elliptic motion of the i^{th} body about the mass-centre of the subsystem of masses $m_1,\ldots,m_{i-1}$. The ε terms provide a measure of the disturbance of this elliptic motion by the other masses. ε^{ki} denotes the disturbance of the i^{th} body by the k^{th} (inferior) body, while $\varepsilon_{\ell i}$ denotes the disturbance of the i^{th} body by the ℓ^{th} (superior) body. These disturbances are normalized with respect to the central "two body" force.

There are several important points to be made concerning the appropriateness of this particular expansion to subsystems of the Solar System. This is a series in powers of α , and assumes that α_{ij} is much less than 1, $\forall_{ij}$, $i < j$, in order to give good convergence. Clearly, as any α approaches 1, some neglected terms become increasingly important and in fact may exceed some of the present terms. While this is not a problem for three-body satellite systems like Earth-Moon-Sun, it does present a problem for planetary systems, where the α's may be quite large.

A second point, following from the first is that the first summation in Equation (16), containing the ε^{ki}s, gives the disturbance on m_i by the inner bodies. These are only the leading terms in the expansion and contain α_{ki}^2 . The second summation, containing the $\varepsilon_{\ell i}$ s, gives the

2.4

disturbance on m_i by the outer bodies. These leading terms contain $\alpha_{\ell i}^3$. Thus there may be neglected terms of order α_{ki}^3 in the first summation which dominate over the leading terms in the second summation.

Walker (1980) recognised these limitations and rederived the equations of motion for planetary systems. Thus the α's were no longer assumed to be small, but m_1 (the Sun) was assumed to be much larger than the other masses (planets). In a method similar to the one described here, he expanded the force function in powers of m_i/m_1, to obtain

$$\ddot{\underset{\sim}{\rho}}_i = G M_i \underset{\sim}{\nabla}_i \left[\frac{1}{\rho_i} \left[1 + \sum_{k=1}^{i-1} \delta^{ki} S'_{ki} + \sum_{\ell=i+1}^{n} \delta_{\ell i} S'_{i\ell} \right] \right] \qquad (i=2,\ldots,n) \tag{18}$$

where

$$\delta^{ki} = \frac{m_k}{m_1} \alpha_{ki}^2 \qquad (2 \leq k < i \leq n) \tag{19}$$

$$\delta_{\ell i} = \frac{m_\ell}{m_1} \alpha_{i\ell}^3 \qquad (2 \leq i < \ell \leq n)$$

and

$$S'_{ij} = \sum_{r=0}^{\infty} \alpha_{ij}^r P_{r+2}(C_{ij}). \tag{20}$$

where $P_r(x)$ is the Legendre polynomial of order r.

Comparing Equations (17) with Equations (19), we see that, correct to the first order in the masses,

$$\delta^{ki} = \varepsilon^{ki} \quad \text{and} \quad \delta_{\ell i} = \varepsilon_{\ell i} \qquad (2 \leq k \leq i \leq \ell \leq n)(k \neq i)(i \neq \ell)$$

using the approximation $M_i \cong m_1$, $\forall i = 1,\ldots,n$, in Equations (17). In addition, if we neglect all but the lowest powers of α_{ij} ratios, then $S'_{ij} = P_2(C_{ij})$, $\forall i,j$ and Equation (18) reduces to Equation (16) correct to 1st order in the masses.

2.5 Empirical Stability Parameters for Three-Body Systems

The major part of this work will be concerned with the three-body problem. It is therefore useful to consider Equations (16) and (17) for n=3 (see Figure 2.3).

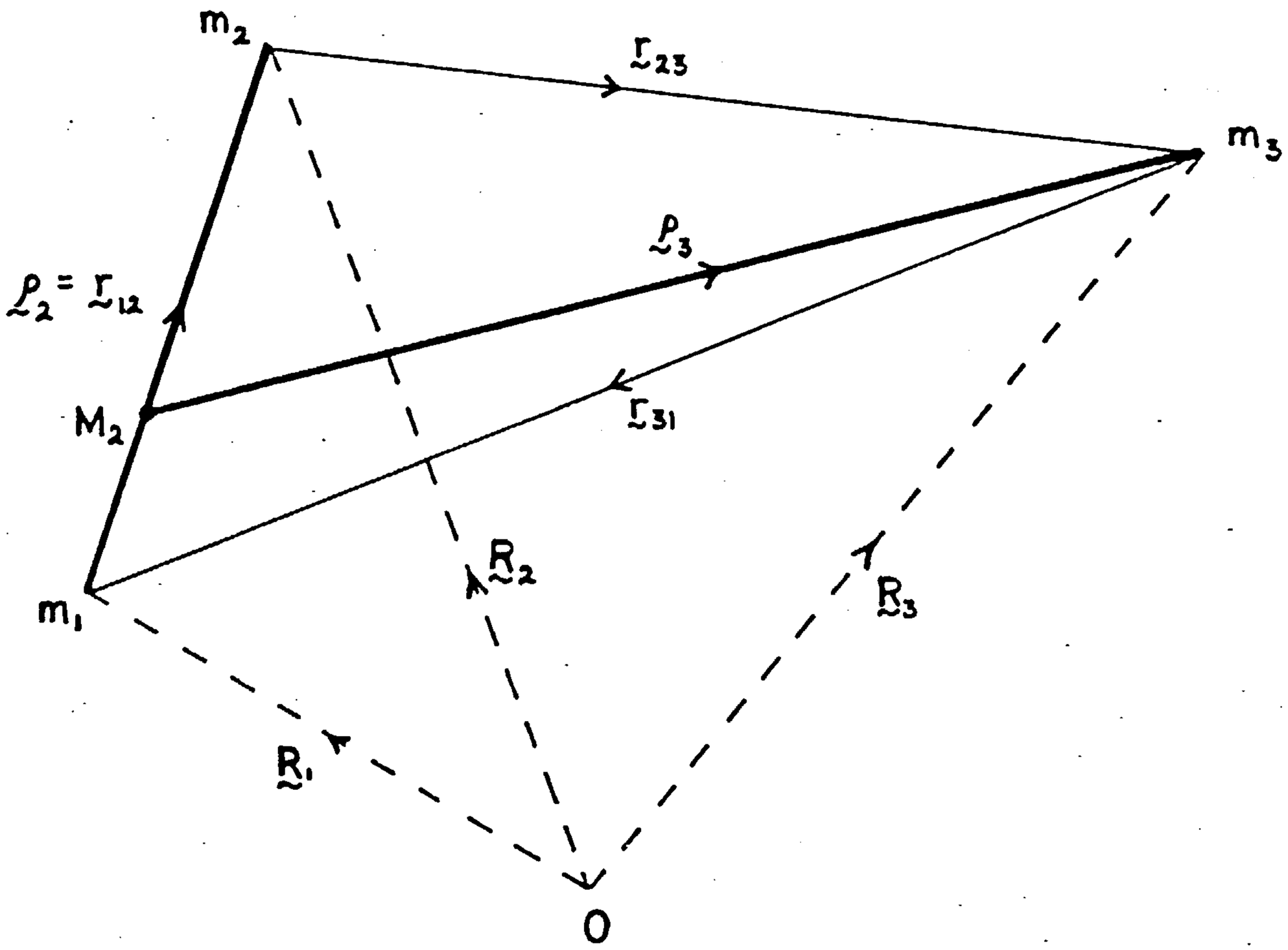


Figure 2.3 Definitions of quantities in the Jacobian Coordinate System in the case n=3.

They become

$$\ddot{\underset{\sim}{\rho}}_2 = G\, M_2\, \underset{\sim}{\nabla}_2 \left[\frac{1}{\rho_2} \left(1 + \varepsilon_{32}\, P_2\, (C_{23}) \right) \right] \tag{21}$$

$$\ddot{\underset{\sim}{\rho}}_3 = G\, M_3\, \underset{\sim}{\nabla}_3 \left[\frac{1}{\rho_3} \left(1 + \varepsilon^{23}\, P_2\, (C_{23}) \right) \right] \tag{22}$$

where

$$\varepsilon^{23} = \frac{m_1\, m_2}{(m_1+m_2)^2}\ \alpha_{23}^2 \tag{23}$$

$$\varepsilon_{32} = \frac{m_3}{m_1 + m_2}\ \alpha_{23}^3 \tag{24}$$

2.5

Walker remarks that the ε^{23} term agrees with the term derived by Brown in his lunar theory (1896).

ε^{23} is a measure of the disturbance of the third body by the other two (in a close binary). This disturbance is due to the fact that m_1 and m_2 are displaced from the centre of mass and have a separation ρ_2. If the separation decreases, α_{23} tends to zero as does ε^{23} and the motion of m_3 tends to perfect Keplerian motion about the centre of mass of m_1 and m_2.

ε_{32} measures the disturbance on the binary (m_1, m_2) by the outer mass m_3. If m_3 recedes from the binary, i.e. ρ_3 increases, then α_{23} (and ε_{32}) tend to zero. The motion of the inner binary becomes Keplerian.

It is seen that for ε^{23}, $\varepsilon_{32} << 1$, we are dealing with two approximately Keplerian orbits. We can assign a set of six osculating elements to each orbit, viz:-

a = semi-major axis of the ellipse

e = eccentricity of the ellipse

i = inclination of the orbital plane to the invariable plane of the system

Ω = longitude of the ascending node of the orbital plane

ω = argument of pericentre

τ = the time of pericentre passage.

Let $(a_2, e_2, i_2, \Omega_2, \omega_2, \tau_2)$ describe the orbit of m_2 about m_1 and $(a_3, e_3, i_3, \Omega_3, \omega_3, \tau_3)$ describe the orbit of m_3 about M_2, the mass-centre of m_1 and m_2. These elements are defined at any instant by the relative positions and velocities of the three bodies, and are linked to ρ_2 and ρ_3 by the following equation,

2.5

$$\rho_2 = \frac{a_2 (1 - e_2^2)}{1 + e_2 \cos f_2}$$

where the position of m_2 in its orbit is given by the true anomaly f_2. There is a similar equation for ρ_3.

If the disturbances are small, i.e. small ε^{23}, ε_{32}, then any periodic changes in the elements should be small, and any secular changes should take place over long periods of time. On the other hand, if either of the ε's approach unity, then the relevant perturbation is of a comparable size to the Keplerian orbit and we should expect violent changes in the orbital elements.

2.6 Summary

In this chapter, the concept of a hierarchical dynamical system has been introduced. We have seen that Jacobian coordinates are ideally suited to modelling simple hierarchical structure. By concentrating on simple hierarchies, we exclude the possibility of studying non-hierarchical systems such as open clusters, and complex hierarchical systems (eg. Castor). In the Solar System, however, the orbits of satellites about planets, and planets about the Sun, can be considered as simple hierarchical systems.

Two expansions of the force function have been performed; one when $\alpha_{ij} << 1$, $\forall i,j=2,\ldots,n$; the other for $m_1 >> m_2,\ldots,m_n$. In each case the leading terms in the series are factored by ε^{ki} and $\varepsilon_{\ell i}$ as given in Equations (17). In most cases, they give a reasonable estimate of the perturbations compared to the central "two-body" forces acting on the Jacobian radius vectors.

2.6

A definition of hierarchical stability was given in Section 2.2. It is the aim in Chapters 3, 4, 5 and 6 to study the stability of hierarchical three-body systems, principally by examining the orbital elements that describe the near-Keplerian motion of the two binary systems, (m_1, m_2) and $(m_1 + m_2, m_3)$.

Many fictitious three-body systems will be examined. They will be grouped not by equal masses, but by equal ε-values, for the following reasons. A system with large masses is not necessarily less stable than a system with smaller masses, provided one of the masses is sufficiently far from the other two. It is the combination of masses and their separations which will determine if a system is stable for all time or if this particular hierarchical structure will only last for a short time before changing irreparably. The ε parameters give an idea of the size of the perturbations on one of the binary systems compared to the central two-body gravitational force. We have already seen that there may be neglected terms in the expansions which may, for some systems, be comparable with the leading terms, or even dominate them. However, we have also seen that the same ε-terms occur in both expansions and this would seem to indicate their usefulness in describing perturbations of all simple hierarchical systems. Systems with common relative perturbations intuitively seem more likely to exhibit the same behaviour than systems with common masses (or mass ratios).

For these reasons, we will examine the stability of three-body systems grouped by equal ε-values. Because it would take too long to examine all possible initial conditions for three-body systems, the examination is restricted to coplanar, initially circular systems that

2.6

always start from conjunction of the three bodies. This limits the choosing of initial parameters to four (ε^{23}, ε_{32}, α_{23}, i_3), having chosen the units of mass, length and time. We set $i_2 = 0^o$ and allow $i_3 = 0^o$ or 180^o corresponding to direct and retrograde systems respectively.

In Chapters 4, 5 and 6 we proceed to investigate and compare the stability of these direct and retrograde systems by means of numerical integration experiments and statistical data reduction. In the next chapter, we review some of the work performed using analytical techniques for determining sufficient conditions for stability.

CHAPTER 3

SUFFICIENT CONDITIONS FOR THE STABILITY OF HIERARCHICAL THREE-BODY SYSTEMS

3.1 Introduction

A definition of hierarchical stability was given in Section 2.2, which is now applied to the general coplanar hierarchical three-body problem. The condition that none of the three bodies escapes is obviously essential for the maintenance of any three-body system. The condition that requires the preservation of the shape and orientation of the orbits is without doubt the most difficult to investigate by numerical experiments and will be discussed at greater length in Chapters 4 and 5. The third condition requires that the ordering of the sizes of $\underset{\sim}{\rho}_2$ and $\underset{\sim}{\rho}_3$ is unchanged, (i.e. $\alpha_{23} < 1$). This is the condition for preserving the particular hierarchical structure of the system and is the subject of this chapter.

In the three-body problem, sufficient conditions can be derived which guarantee the preservation of a hierarchy for all subsequent times, by examining the topology of the regions of forbidden motion of the three bodies. This is analogous to the Hill stability criterion for the restricted three-body problem (Section 1.2). The treatment given here is by Marchal and Saari (Section 3.2). It holds for the non-coplanar three-body problem and indicates with some clarity why the conditions given are sufficient but not necessary for hierarchical preservation. Zare and Szebehely derived similar results for the coplanar three-body problem.

We saw in Chapter 2, that for a system with prescribed masses, the perturbations on the orbits decrease as α_{23} decreases. Thus, in all likelihood, a system will be more stable, the smaller we make α_{23}. Indeed we may find the critical value of α_{23}, called α_c, below which a system will have its hierarchical structure preserved for all time. In Section 2.4 the method for determining α_c as a function of ε^{23}, ε_{32} is

3.1

described and results for direct and retrograde coplanar three-body systems are given.

3.2 A Review of Work by Marchal and Saari

The theory presented here is described in much greater detail by Marchal and Saari (1975) and Marchal (1985). In the circular restricted three-body problem, there exist regions of forbidden motion bounded by the zero velocity curves which are determined by the value of the Jacobi constant, (Section 1.2). If these regions enclose two of the bodies with the third outside, then the hierarchical nature is preserved for all time. There is an analogous result for the general three-body problem.

The constant of angular momentum $\underset{\sim}{c}$, and the total energy of the system h, determine manifolds in the phase space of the system. By projecting these manifolds into the configuration space we find there are similar regions of forbidden motion. As in the restricted case, for particular values of $\underset{\sim}{c}$ and h, these regions may enclose the inner binary for all time, thus preserving the hierarchy.

Consider a three-body system with masses m_1, m_2, m_3. We will use the notation of Chapter 2, but in addition define the following quantities,

$$M = m_1 + m_2 + m_3 \tag{1}$$

$$M^* = m_1m_2 + m_1m_3 + m_2m_3 \tag{2}$$

$$a = - GM^*/2h \tag{3}$$

$$p = Mc^2/G(M^*)^2. \tag{4}$$

a and p are respectively the semi-major axis and the semi-latus rectum of the elliptic orbit for two of the bodies when the third has negligible mass (i.e. the restricted three-body problem). Define the mean quadratic

3.2

distance* σ and the mean harmonic distance ν by

$$M^*\sigma^2 = m_1m_2r_{12}^2 + m_1m_3r_{13}^2 + m_2m_3r_{23}^2 \tag{5}$$

$$\frac{M^*}{\nu} = \frac{m_1m_2}{r_{12}} + \frac{m_1m_3}{r_{13}} + \frac{m_2m_3}{r_{23}} \quad . \tag{6}$$

Defining the moment of inertia as 2I and recalling the force function U from Equation 2.2, it can be shown that

$$M^*\sigma^2 = 2MI \tag{7}$$

$$M^*/\nu = U/G \tag{8}$$

Define the Sundman function j,

$$j = \frac{\bar{p}}{2\sigma} + \frac{\sigma}{2a} + \frac{\sigma\,\sigma'^2}{2\,GM} \tag{9}$$

where $\sigma' = d\sigma/dt$. Using Equations 1-9, the following classical results may be rewritten.

(a) The Lagrange-Jacobi identity

$$\frac{d^2I}{dt^2} = U + 2h$$

becomes

$$\frac{d^2(\sigma^2)}{dt^2} = GM\left(\frac{1}{\nu} - \frac{1}{a}\right) \tag{10}$$

(b) The Sundman inequality

$$4I(U+h) \geq \left(\frac{dI}{dt}\right)^2 + c^2$$

becomes

$$\frac{\sigma}{\nu} \geq j \tag{11}$$

---

*Marchal and Saari denote the mean quadratic distance by ρ. To avoid ambiguity with the Jacobian radius vectors $\underset{\sim}{\rho}$, we shall denote it by σ.

3.2

It can be shown that

$$\min(r_{12}, r_{13}, r_{23}) \leq \nu \leq \sigma \leq \max (r_{12}, r_{13}, r_{23}).$$

It is easily seen that $\nu = \sigma$ in the restricted problem or when all three bodies form an equilateral triangle.

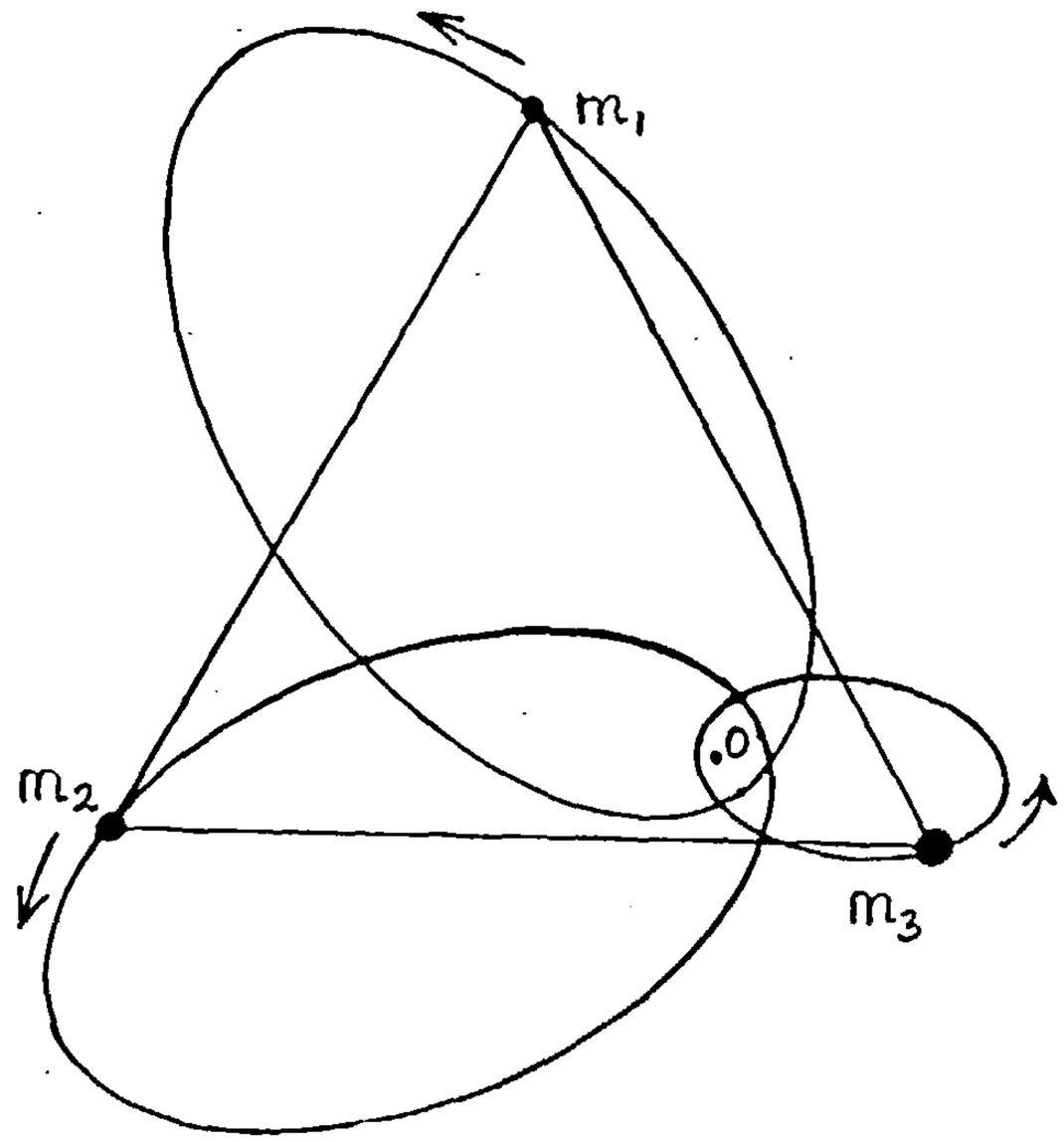
This is one of two classical sets of solutions, given by Lagrange and Euler. In these solutions, the configuration of the system does not alter with time, although the scale might. The acceleration on each of the bodies is always radial from the centre of mass of the system and is proportional to the distance of the body from the mass centre.

The Lagrangian configuration mentioned above is shown in Figure 3.1(a). For this solution, $r_{12} = r_{13} = r_{23} = \sigma = \nu$, and $j = 1$, for all time. Thus the three bodies always form an equilateral triangle by describing coplanar ellipses with equal eccentricities, about the mass centre O.

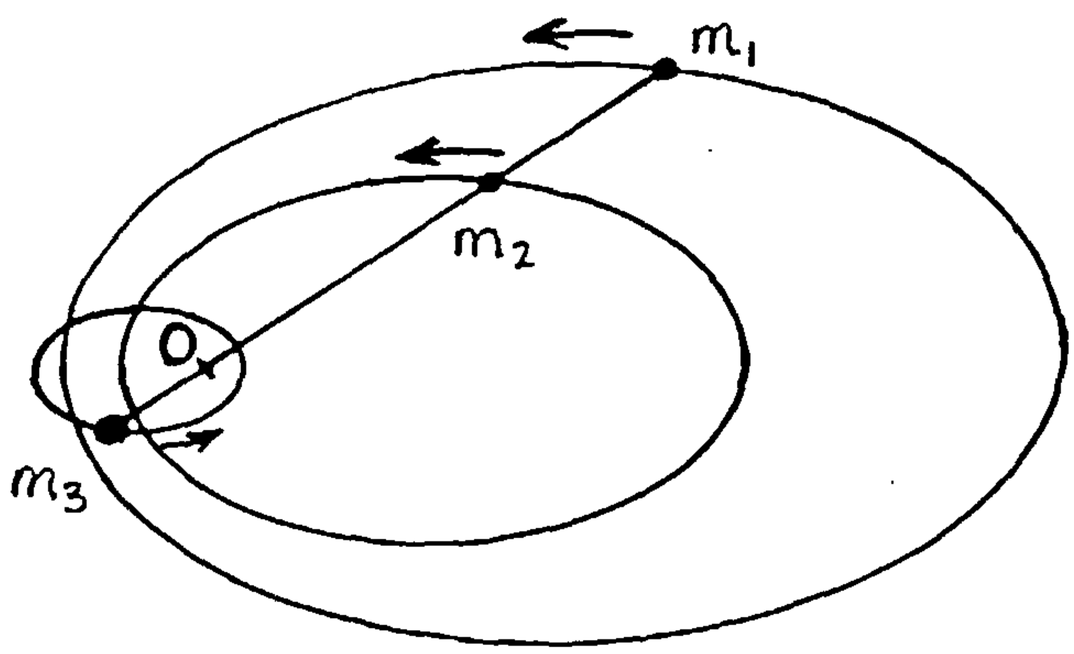
The Eulerian solutions (Figure 3.1(b)) are given when the bodies are arranged in a straight line for all time by describing homothetic ellipses about the mass centre. There are three such solutions depending on whether the middle body is m_1,m_2 or m_3. In each case σ/ν and j are both constant but not equal to 1, as in the Lagrangian solutions. The conditions on accelerations are automatically satisfied for the Lagrangian solutions. For the Eulerian solution, this condition defines the position of one of the bodies with respect to the other two, which can be derived in general by solving a quintic polynomial. This is more fully discussed in the next section.

The ratio σ/ν , defined by Equations (2), (5), (6), is dependent on the shape, but not on the scale of the triangle described by the three bodies. Figures 3.2 show the contours of σ/ν in the configuration space

(a)



(b)



Figures 3.1: Examples of (a) Lagrangian motion, (b) Eulerian motion in the general three-body problem, with $m_1 < m_2 < m_3$.

3.2

for masses 10, 5 and 2. In each diagram the configuration is described by the position of one of the bodies relative to the other two. All the diagrams are equivalent but, as we will see, all are useful for describing the preservation of the hierarchy.

In all cases, σ/ν attains its minimum (unity) at the triangular Lagrange points L_4 and L_5, and tends to infinity at infinity and the two fixed masses. There are saddle points at the Euler points L_1, L_2, L_3.

Throughout this section we will order the masses in increasing size as in Figures 3.2, i.e.

$$0 < m_1 \leqslant m_2 \leqslant m_3 \ . \tag{12}$$

We denote each of the Eulerian configurations by the points L_1, L_2, L_3, where L_i is the configuration which has m_i as the middle body. By studying Figures 3.2, we see that for the given ordering of the masses (Inequalities (12)), we have

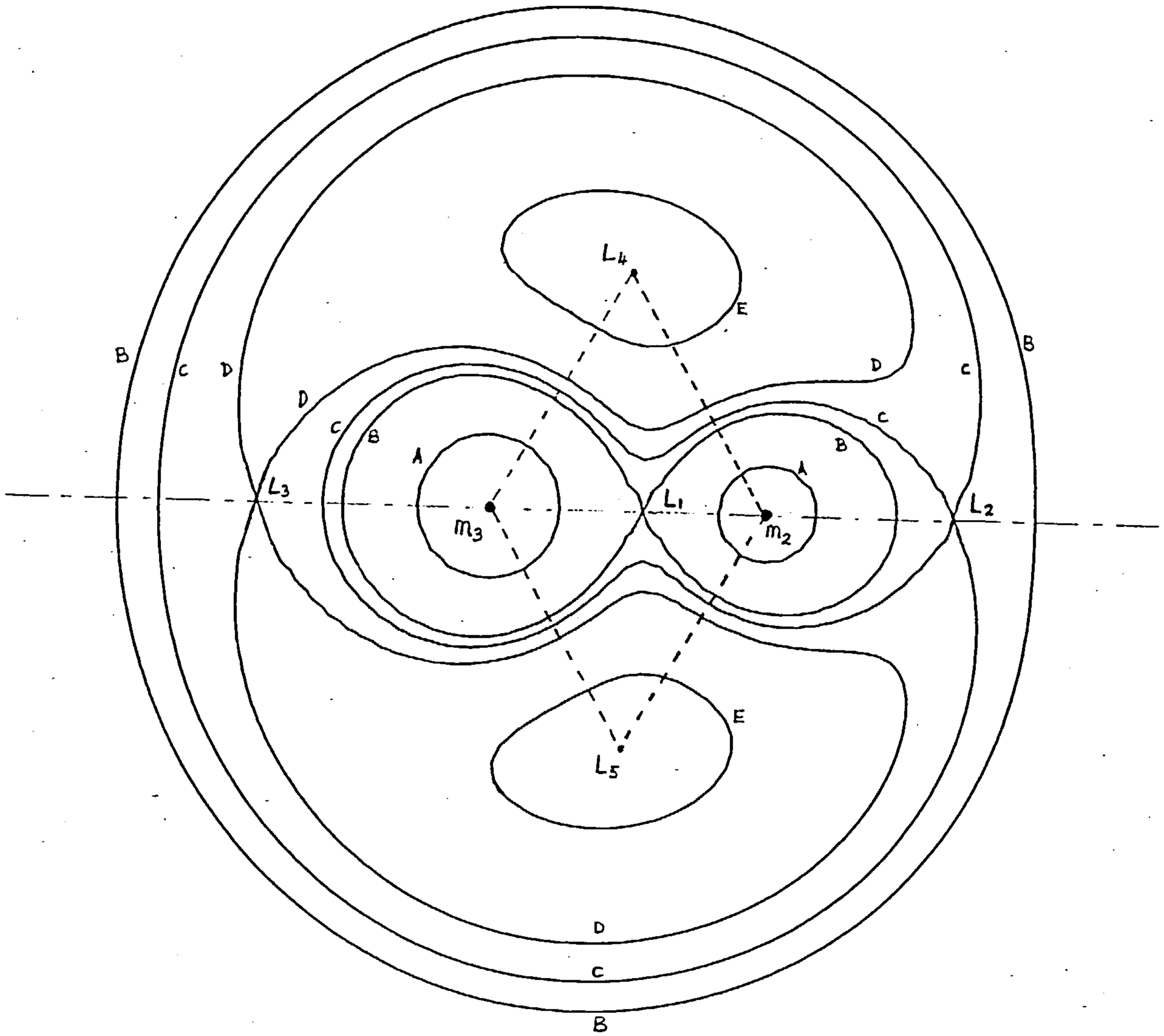
$$\frac{\sigma}{\nu}(L_1) \geqslant \frac{\sigma}{\nu}(L_2) \geqslant \frac{\sigma}{\nu}(L_3) > 1$$

Consider Equation (9) and the Sundman Inequality, i.e.

$$\frac{\sigma}{\nu} \geqslant j = \frac{p}{2\sigma} + \frac{\sigma}{2a} + \frac{\sigma\,\sigma'^2}{2\,GM} \tag{13}$$

When the total energy of the system is positive, the system is deemed immediately unstable, since one of the bodies is in the process of escape. Assume therefore that $h < 0$, hence $a > 0$. For a system with prescribed masses, a and p, the minimum of the right hand side of Equation (13), with respect to σ, is obtained when $\sigma' = 0$ and $\sigma = \sqrt{ap}$, hence

$$h < 0 \Rightarrow \frac{\sigma}{\nu} \geqslant \sqrt{\frac{p}{a}} \tag{14}$$



Figures 3.2: σ/ν contours for $m_1 = 2$, $m_2 = 5$ and $m_3 = 10$, (in terms of the position of one body with respect to the other two). The values of σ/ν for the contours are: (A) 1.5, (B) 1.1562, (C) 1.1283, (D) 1.0952, (E) 1.02. $\sigma/\nu = 1$ at L_4, L_5 and is infinite at the two fixed masses and infinity. All three diagrams are equivalent.
(a) m_3, m_2 are fixed and the position of m_1 is varied.

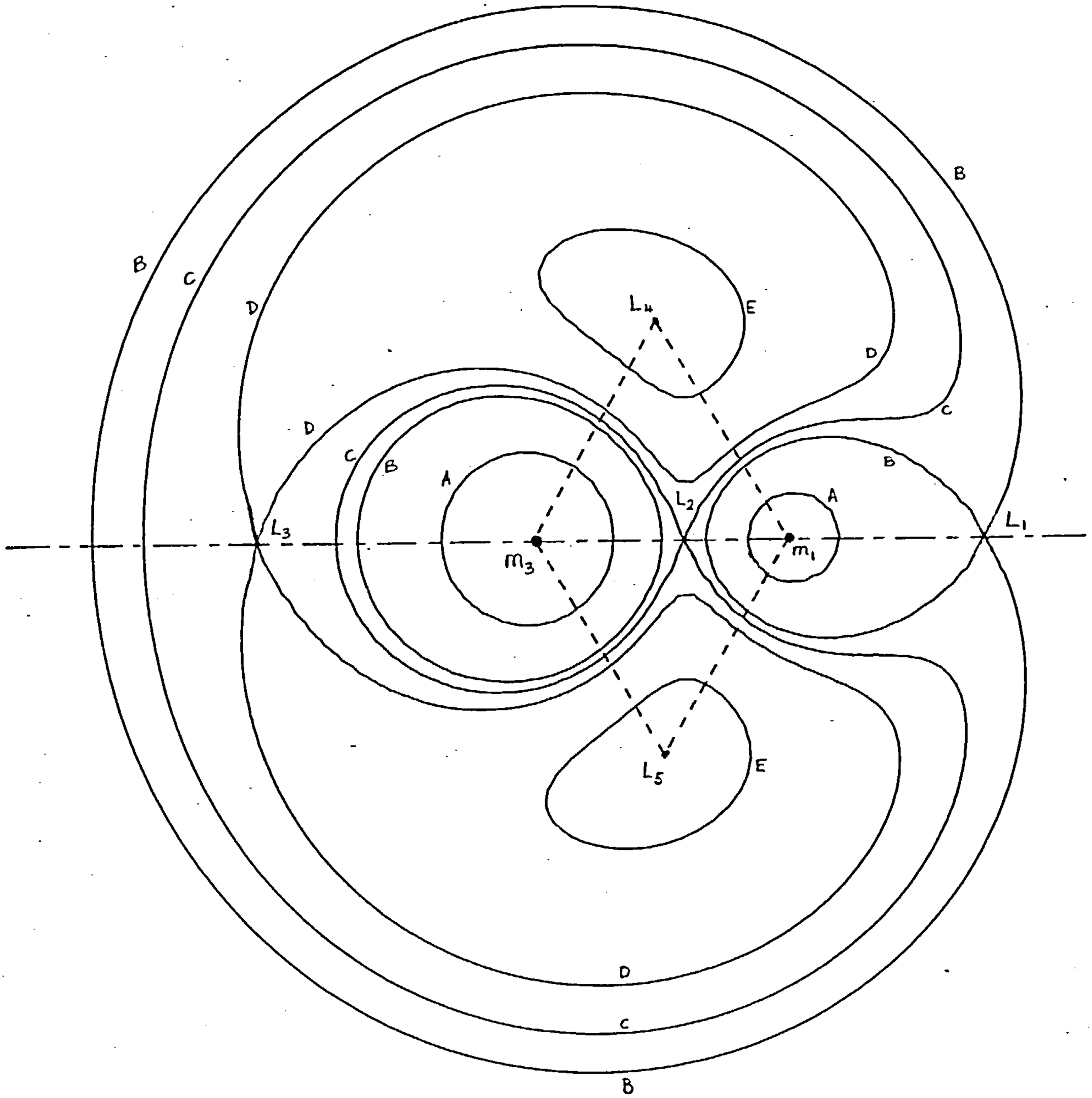


Figure 3.2(b): σ/ν contours when m_3, m_1 are fixed and the position of m_2 is varied.

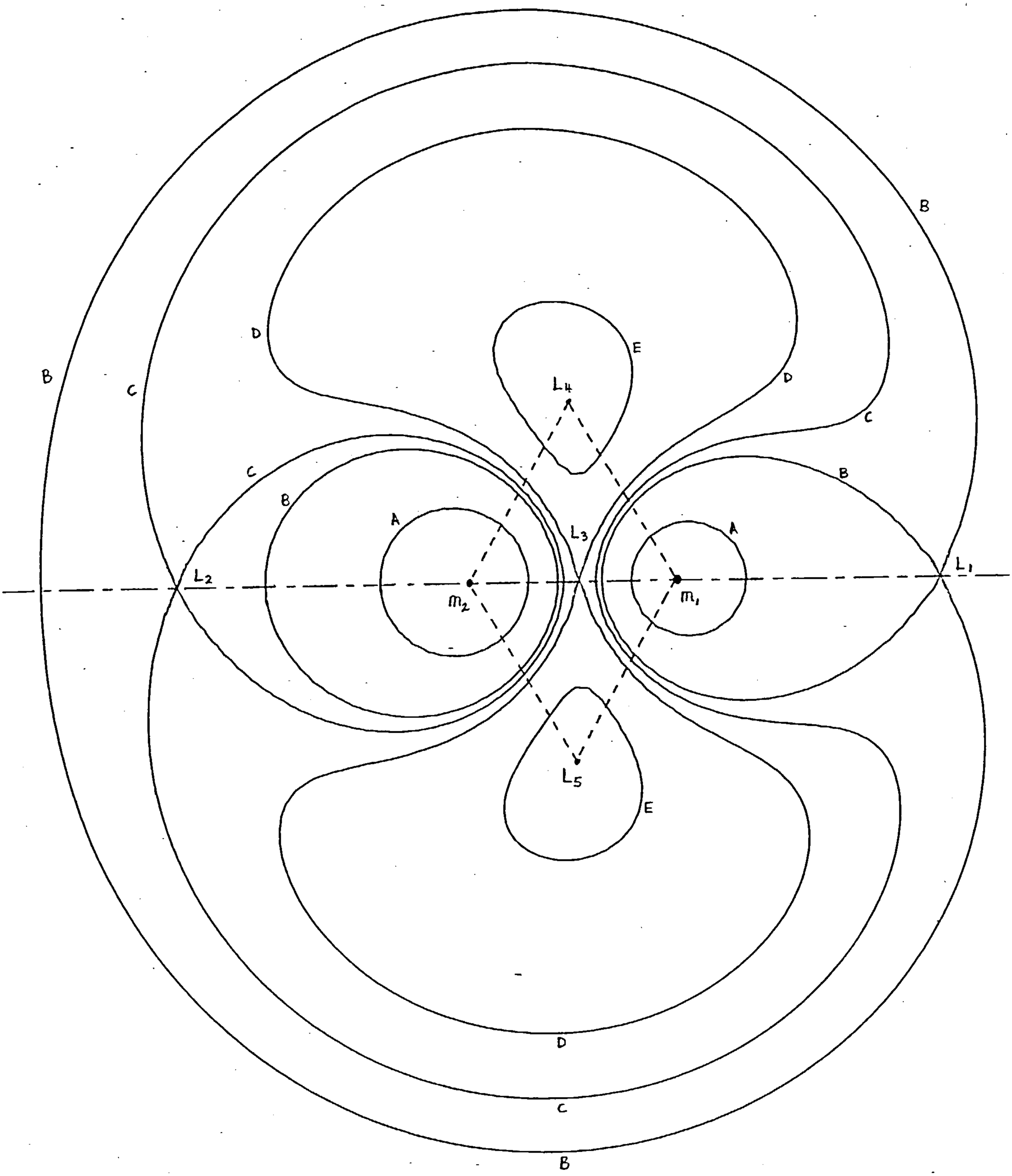


Figure 3.2(c): σ/ν contours when m_2, m_1 are fixed and the position of m_3 is varied.

3.2

where $\sqrt{p/a}$ is the minimum of the Sundman function, and

$$\frac{p}{a} = \frac{-2M}{G^2(M^*)^3} \cdot c^2h \qquad (15)$$

Since $\sigma \geq \nu$, Inequality (14) gives no restrictions on the allowed configurations of the system, so long as $p/a < 1$, i.e. Inequality (14) is automatically satisfied for all configurations. If p/a is slightly greater than one, then there are small forbidden regions around the L_4 and L_5 points, since σ/ν cannot tend to unity. The larger the value of p/a is, the more extensive are the forbidden regions.

The analogy with the zero velocity curves in the restricted three-body problem now becomes apparent, as similar forbidden regions appear in both problems. They are characterised in the restricted problem by the Jacobi constant and in the general problem by c^2h. We can thus apply Hill-type stability to the general problem by considering the topology of the forbidden regions.

If $\sqrt{p/a} > \sigma/\nu\ (L_1)$, then the three collinear Eulerian configurations are forbidden and there are three unconnected zones of allowed motion. Thus one body is necessarily isolated from the other two, i.e. any hierarchical arrangement is preserved for all time. If $\sqrt{p/a} = \sigma/\nu(L_1)$, then the L_1 point is an allowed configuration, but no body can pass from one region, through this neck, to another.

If $\sigma/\nu(L_1) > \sqrt{p/a} > \sigma/\nu(L_2)$, then the L_1 point is an allowed configuration and there are now only two unconnected regions of possible motion, one of finite extent, the other of infinite extent. If the hierarchical arrangement is such that a binary exists in the finite region, with the third mass in the infinite region, then the hierarchy will still be preserved. If this is not the case, then there is a possibility

3.2

that the ordering of the Jacobian radius vectors may change and the hierarchy will be broken up.

If $\sqrt{p/a} = \sigma/\nu(L_2)$, then the L_2 point is an allowed configuration, but no exchange of bodies through this neck is possible, so the results of the previous paragraph still apply.

If $\sqrt{p/a} < \sigma/\nu(L_2)$, then the L_1 and L_2 points are allowed configurations. There is now one region of possible motion and thus there is no guarantee that any hierarchy will be preserved. If in addition $\sqrt{p/a} < \sigma/\nu(L_3)$ then there are two unconnected regions of forbidden motion around each of the L_4, L_5 points.

By way of example consider the system described in Figures 3.2, where $m_1 = 2$, $m_2 = 5$, $m_3 = 10$. Since m_1 is the smallest mass, Figure 3.2(a), is the one we are most familiar with. On comparing it with the zero velocity curves in the restricted problem, we see that the forbidden regions open up in the same order with L_1 in between the two fixed masses, as usual. As m_1 tends to zero, the general problem tends to the restricted problem, so this is to be expected. If however we consider Figures 3.2(b),(c), then the smallest mass is no longer the one that is allowed to move over the configuration space. Thus the order of the L_1, L_2, L_3 points, and hence of connection of the allowed regions is not so familiar. It is however equivalent to the order in Figure 3.2(a).

We are interested in determining the critical values of $\sqrt{p/a}$ for which the preservation of the hierarchy is guaranteed for all time. We now impose a hierarchical structure on the three body system, with one body significantly further away from the other two, which form a binary. There are three different hierarchies, depending on the choice of m_1, m_2 or m_3 for the outer body.

3.2

Let m_1 be the outer body. In addition, let $\sqrt{p/a} \gtrsim \sigma/\nu(L_1)$ (e.g. contour A, $\sigma/\nu = 1.5$). In Figure 3.2(a), m_1 must lie in the unbounded region of allowed motion. In Figure 3.2(b), m_2 must lie in the region around m_3. In Figure 3.2(c), m_3 must lie in the region around m_2. This is in order to conform to the imposed hierarchy. Since $\sqrt{p/a} \gtrsim \sigma/\nu(L_1)$, the hierarchy is guaranteed for all time.

Let $\sigma/\nu(L_1) > \sqrt{p/a} \gtrsim \sigma/\nu(L_2)$. In Figure 3.2(a), the regions around m_3 and m_2 have been connected, but m_1 is still unable to approach either through the L_2 or L_3 points. The hierarchy is still preserved. In Figures 3.2(b),(c), m_1 is now free to move into the unbounded region but still cannot approach either m_2 or m_3, which are in the other region.

Let $\sqrt{p/a} < \sigma/\nu(L_2)$. m_1 can now approach m_2 or m_3 through the "L_2 neck". There is no guarantee of hierarchical preservation. Thus the critical value of $\sqrt{p/a}$ is $\sigma/\nu(L_2)$.

Now consider the case when m_2 is the outer body and $\sqrt{p/a} \gtrsim \sigma/\nu(L_1)$. In Figure 3.2(a), m_1 must lie in the region around m_3. In Figure 3.2(b), m_2 must lie in the unbounded region. In Figure 3.2(c), m_3 must lie in the region around m_1. The given hierarchical arrangement is guaranteed for all time.

Let $\sqrt{p/a} < \sigma/\nu(L_1)$. In Figure 3.2(a), m_1 is free to approach m_2, via the L_1 neck, changing the hierarchy. In Figure 3.2(b), m_2 is free to approach m_1, through the L_1 neck. In Figure 3.2(c), m_3 is able to wander into the unbounded region and move far from m_1, compared to m_2. In each case, there is a possible alteration of the hierarchy of the system. Thus the critical value of $\sqrt{p/a}$ is $\sigma/\nu(L_1)$.

When m_3 is the outer body, the results are similar to the case when m_2 is outermost. The critical value of $\sqrt{p/a}$ is $\sigma/\nu(L_1)$.

It is seen that for a hierarchical three-body system, the critical value of $\sqrt{p/a}$ for hierarchical preservation is $\sigma/\nu(L_2)$ if the smallest

3.2

mass is outermost. Otherwise the critical value is $\sigma/\nu(L_1)$. By considering Figures 3.2, it is clear that in any case, the critical value of σ/ν is derived when the three bodies are in the Euler configuration which has the smaller of the two masses in the close binary, as the central mass. If the hierarchy is $((m_3,m_2),m_1)$ then the relevant configuration is $m_3 - m_2 - m_1$ (L_2 point); whereas for hierarchies $((m_2,m_1),m_3)$ and $((m_1,m_3),m_2)$, the relevant configuration is $m_2-m_1-m_3$ (L_1point).

In Chapter 2, it was pointed out that a hierarchical system of three bodies could be described by two binary orbits with osculating elements. In the next section we shall use this theory to derive critical values of $\alpha_{23} = \rho_2/\rho_3$, such that any system with α_{23} less than the critical value will be guaranteed as hierarchically preserved, for all time.

3.3 Determination of Critical Surfaces

The results of the previous section were also derived by Zare (1976, 1977) for the coplanar three-body problem by considering the surfaces of zero velocity of the reduced Hamiltonian. He went on to derive a quintic polynomial which has as a solution, one of the Eulerian configurations, from which the critical value of c^2h, could be obtained. Walker and Roy apply the analytical stability criterion of Zare to hierarchical coplanar three-body systems, whose binary systems (m_1,m_2), $((m_1+m_2),m_3)$ have circular orbits initially. They numerically derive the critical value of α_{23} as a function of $\epsilon^{23},\epsilon_{32}$, for which any system, with α_{23} less than this critical value, is guaranteed to maintain its present hierarchy. This approach will now be summarised in terms of the notation of Marchal and Saari.

3.3

Throughout the remainder of this chapter as well as Chapters 4, 5, 6, we will drop the subscript '23' on α_{23} and merely denote the ratio ρ_2/ρ_3 by α . There will be no ambiguity as these chapters are solely concerned with the three-body problem.

Using the notation of Chapter 2, we define the hierarchy of a three-body system in terms of two binary orbits, exhibiting approximate Keplerian motion. The orbits are m_2 about m_1, and m_3 about M_2, the centre of mass of m_1, m_2. These orbits are characterised by the osculating elements a_2, $e_2, \tilde{\omega}_2, \tau_2$ and a_3, $e_3, \tilde{\omega}_3$, τ_3, respectively, assuming the orbits are coplanar.

The total energy of the system is approximately

$$h = \frac{-G}{2}\left[\frac{m_1 m_2}{a_2} + \frac{M_2 m_3}{a_3}\right] \qquad (16)$$

being the sum of the individual Keplerian energies. It does neglect the displacement of m_1 and m_2 from M_2 which gives rise to a small error in the energy of the outer binary for sufficiently small α .

The angular momentum of the system is exactly the sum of the angular momenta of the binaries.

$$c = \sqrt{G}\left[m_1 m_2 \left[\frac{a_2(1-e_2^2)}{M_2}\right]^{\frac{1}{2}} \pm M_2 m_3 \left[\frac{a_3(1-e_3^2)}{M_3}\right]^{\frac{1}{2}}\right] \qquad (17)$$

where the + sign refers to direct systems and the - sign refers to retrograde systems. Rewriting Equation (15) in the notation of Chapter 2, we have

$$\frac{p}{a} = -\frac{2\,M_3}{G^2(M^*)^3} \cdot c^2 h \qquad (18)$$

3.3

where $M^* = m_1m_2 + m_1m_3 + m_2m_3$ as before. Combining Equations (16), (17), (18),

$$\frac{P}{a} = \frac{M_3}{(M^*)^3}\left[m_1^2m_2^2m_3\ (1-e_2^2)\ \frac{a_2}{a_3} \pm 2m_1m_2m_3^2\ M_2\ \left(\frac{M_2}{M_3}\right)^{\frac{1}{2}}((1-e_2^2)(1-e_3^2))^{\frac{1}{2}}\left(\frac{a_2}{a_3}\right)^{\frac{1}{2}}\right.$$

$$+ \frac{m_1^3m_2^3}{M_2}\ (1-e_2^2) + \frac{m_3^3\ M_2^3}{M_3}\ (1-e_3^2)$$

$$\left. \pm 2m_1^2m_2^2m_3\ \left(\frac{M_2}{M_3}\right)^{\frac{1}{2}}((1-e_2^2)(1-e_3^2))^{\frac{1}{2}}\left(\frac{a_3}{a_2}\right)^{\frac{1}{2}} + \frac{m_1m_2m_3^2\ M_2^2}{M_3}\ (1-e_3^2)^{\frac{1}{2}}\left(\frac{a_3}{a_2}\right)\right] \tag{19}$$

We now normalise the masses with respect to the mass of the inner binary, M_2, by defining

$$\left.\begin{aligned} \mu &= \frac{m_2}{M_2} \\ \mu_3 &= \frac{m_3}{M_2} \end{aligned}\right\} \qquad \mu^* = \mu(1-\mu) + \mu\mu_3 + (1-\mu)\mu_3 \tag{20}$$

Hence, Equation (19) becomes

$$\frac{P}{a} = \frac{(1+\mu_3)}{(\mu^*)^3}\left[(1-\mu)^2\mu^2\mu_3(1-e_2^2)\ \frac{a_2}{a_3} \pm 2\frac{(1-\mu)\ \mu\ \mu_3^2}{(1+\mu_3)^{\frac{1}{2}}}\ ((1-e_2^2)(1-e_3^2))^{\frac{1}{2}}\left(\frac{a_2}{a_3}\right)^{\frac{1}{2}}\right.$$

$$+ (1-\mu)^3\mu^3(1-e_2^2) + \frac{\mu_3^3}{1+\mu_3}\ (1-e_3^2)$$

$$\left. \pm 2(1-\mu)^2\frac{\mu^2\mu_3}{(1+\mu_3)^{\frac{1}{2}}}\ ((1-e_2^2)(1-e_3^2))^{\frac{1}{2}}\ \left(\frac{a_2}{a_3}\right)^{-\frac{1}{2}} + \frac{(1-\mu)\mu\mu_3^2}{1+\mu_3}\ (1-e_3^2)\left(\frac{a_2}{a_3}\right)^{-1}\right] \tag{21}$$

Without loss of generality we may assume that $m_2 \leq m_1$, i.e. $\mu \leq \frac{1}{2}$. μ_3 may of course take any positive value. Through Equation (21), the

3.3

mass ratios, the shape and relative sizes of the binary orbits, define a value of p/a. For a guarantee of hierarchical preservation, we require that $\sqrt{p/a} \gtrsim \sigma/\nu$, where σ/ν is evaluated at the Eulerian configuration with m_2 as the middle mass.

In order to find this value of σ/ν, we normalise the masses as before and also normalise the distances r_{ij} with respect to r_{12}. We choose 2-dimensional coordinates such that m_1 is at the origin, m_2 at (1,0), m_3 at (1 + x,y). Thus the configuration is described by x,y. Thus,

$$\left.\begin{aligned} r_{12} &= 1 \\ r_{13} &= ((1+x)^2 + y^2)^{\frac{1}{2}} \\ r_{23} &= (x^2 + y^2)^{\frac{1}{2}} \end{aligned}\right\} \qquad (22)$$

After normalising the masses,

$$\left(\frac{\sigma}{\nu}\right)^2 (\mu^*)^3 = f^2 g. \qquad (23)$$

where

$$f = \mu(1-\mu) + \frac{(1-\mu)\mu_3}{r_{13}} + \frac{\mu\mu_3}{r_{23}} \qquad (24)$$

$$g = \mu(1-\mu) + (1-\mu)\,\mu_3\, r_{13}^2 + \mu\mu_3\, r_{23}^2 \qquad (25)$$

The critical values of $f^2 g$ are given by

$$\partial(f^2 g)/\partial x = \partial(f^2 g)/\partial y = 0.$$

Noting that $f > 0$ we are required to solve

$$f\frac{\partial g}{\partial x} + 2g\,\frac{\partial f}{\partial x} = 0 \qquad (26)$$

$$f\frac{\partial g}{\partial y} + 2g\,\frac{\partial f}{\partial y} = 0\ . \qquad (27)$$

The rest is straightforward, if rather tedious. It can be verified that the Lagrangian solution, $r_{13} = r_{23} = 1$, satisfies Equations (26) and (27). It is also easy to show that $y = 0$ is a solution of Equation (27).

3.3

Substitution into Equation (26) yields three solutions for x in the domains $(-\infty, -1)$, $(-1,0)$, $(0,\infty)$ respectively. Since we are interested in the solution when m_2 is the middle mass, we take the positive value of x. Hence Equation (26) becomes

$$p(x) = x^5 + (3-\mu)x^4 + (3-2\mu)x^3 - (\mu+3\mu_3)x^2 - (2\mu+ 3\mu_3)x - (\mu+\mu_3) = 0 \tag{28}$$

which by Descartes Rule of Signs, has only one real positive solution.

We shall be concerned only with coplanar systems that have initially circular orbits. Thus $e_2 = e_3 = 0$, and $\alpha = \rho_2/\rho_3 = a_2/a_3$ at time zero (although not necessarily at subsequent times). This means that to every pair of mass ratios, we may ascribe α_c, a critical value of α s.t. any systems that have an initial α less than α_c, are guaranteed to preserve the hierarchy for all time. If α exceeds α_c, no such guarantee exists.

The method is as follows. For prescribed μ,μ_3, calculate x from Equation (28), by a suitable iterative method. For example, a Newton-Raphson method can be used, where

$$x_2 = x_1 - \frac{p(x_1)}{p'(x_1)}$$

is the new approximation of x, given a previous approximation x_1.

From the value of x, f^2g can be calculated from Equation (24) and (25), with $r_{13} = 1 + x$, $r_{23} = x$. In order to find α_c, we must set $(\sigma/\nu)^2 = p/a$, with $\alpha = \alpha_c$, $e_2 = e_3 = 0$. Hence, by Equations (21) and (23),

$$\begin{aligned} f^2g = {} & (1-\mu)^2\mu^2\mu_3(1+\mu_3)\alpha_c \pm 2(1-\mu)\mu\mu_3^2(1+\mu_3)^{\frac{1}{2}}\ \alpha_c^{\frac{1}{2}} \\ & +(1-\mu)^3\mu^3(1+\mu_3)+\mu_3^3 \pm 2(1-\mu)^2\mu^2\mu_3(1+\mu_3)^{\frac{1}{2}}\ \alpha_c^{-\frac{1}{2}} \\ & + (1-\mu)\mu\mu_3^2\ \alpha_c^{-1} \end{aligned} \tag{29}$$

3.3

Equation (29) may be expressed as a quartic polynomial in $\alpha_c^{\frac{1}{2}}$, hence α_c may be described by an iterative scheme similar to that used for deriving x.

Figures 3.3 and 3.4 give the surfaces $\alpha_c = \alpha_c(\mu, \mu_3)$ for direct and retrograde systems. In each case the range is $10^{-9} \leq \mu \leq 10^{-0.5}$, $10^{-9} \leq \mu_3 \leq 10^{7}$, with logarithmic scales in μ, μ_3. For the direct systems, the general behaviour is as expected. α_c increases monotonically to unity as μ and μ_3 tend to zero. This is not the case for the retrograde systems. Indeed, for μ much greater than μ_3, α_c is almost zero. This is better seen in Figure 3.5, which shows α_c against μ_3 for $\mu = 0.5$, the largest value of μ allowed. We see that for $\mu_3 << 1$, $\alpha_c \sim \mu_3^2$. Thus any retrograde system with small μ_3 will in all probability have no guarantee of hierarchical preservation. This seems counter-intuitive, and will be discussed again later.

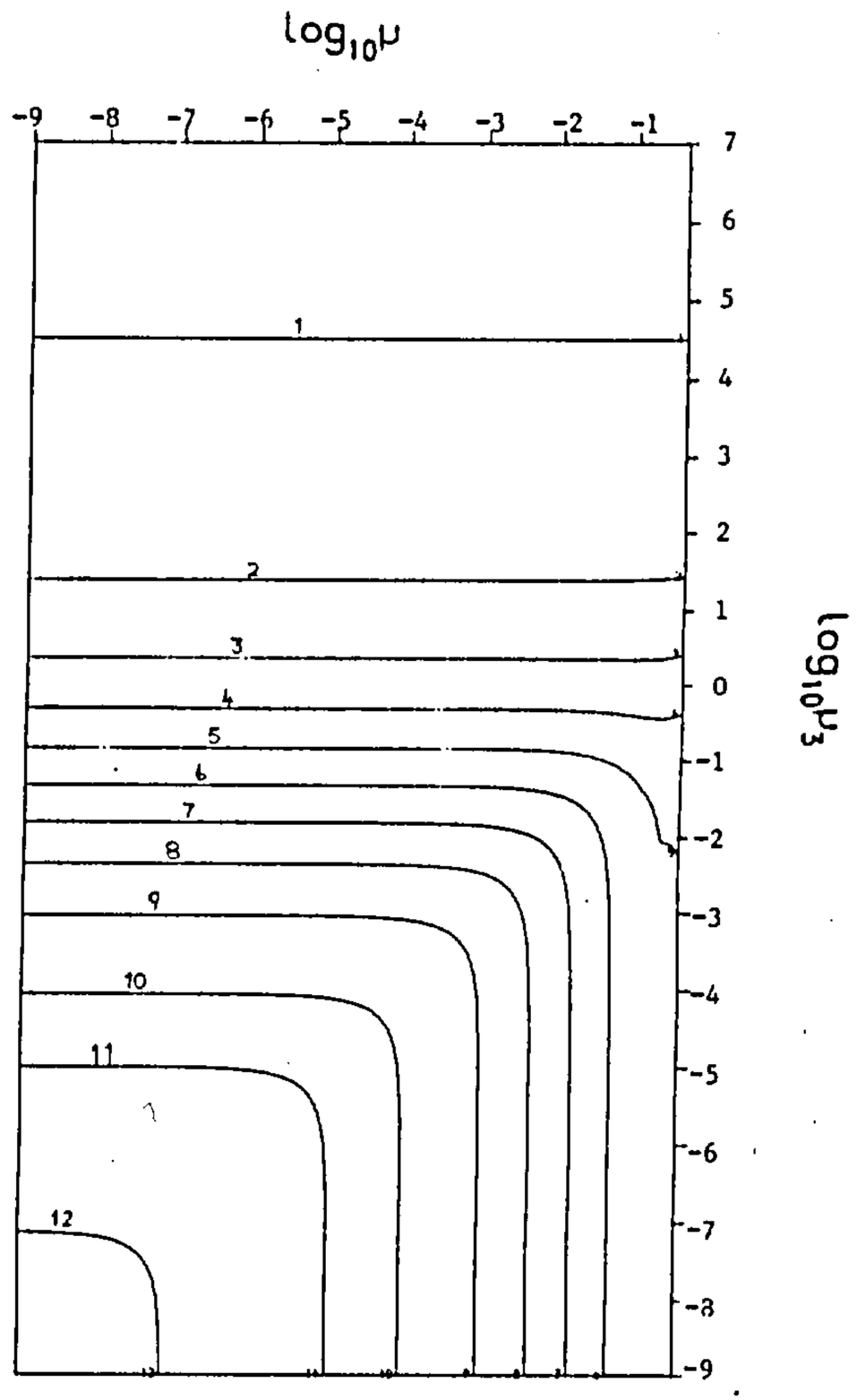
It is also of interest to see the surfaces $\alpha_c = \alpha_c(\varepsilon^{23}, \varepsilon_{32})$. The easiest method is to calculate from the (μ, μ_3, α_c) data, the corresponding values of $\varepsilon^{23}, \varepsilon_{32}$, from

$$\varepsilon^{23} = \mu(1-\mu)\alpha^2 \qquad (30)$$

$$\varepsilon_{32} = \mu_3 \alpha^3 \qquad (31)$$

where α is set equal to α_c. By the use of various numerical interpolation routines, approximate contours may be drawn.

Figures 3.6 and 3.7 show for direct and retrograde systems respectively, how a grid of (μ, μ_3) values transform into the $(\varepsilon^{23}, \varepsilon_{32})$ space when $\alpha = \alpha_c$. The scales are logarithmic with $\mu = 10^p$, $p = -9, -8.5, -8, \ldots, -0.5$, and $\mu_3 = 10^q$, $q = -9, -8.5, \ldots, 6.5, 7$. The arrows indicate increasing μ with μ_3 fixed, or vice versa. There is a region in $(\varepsilon^{23}, \varepsilon_{32})$ space



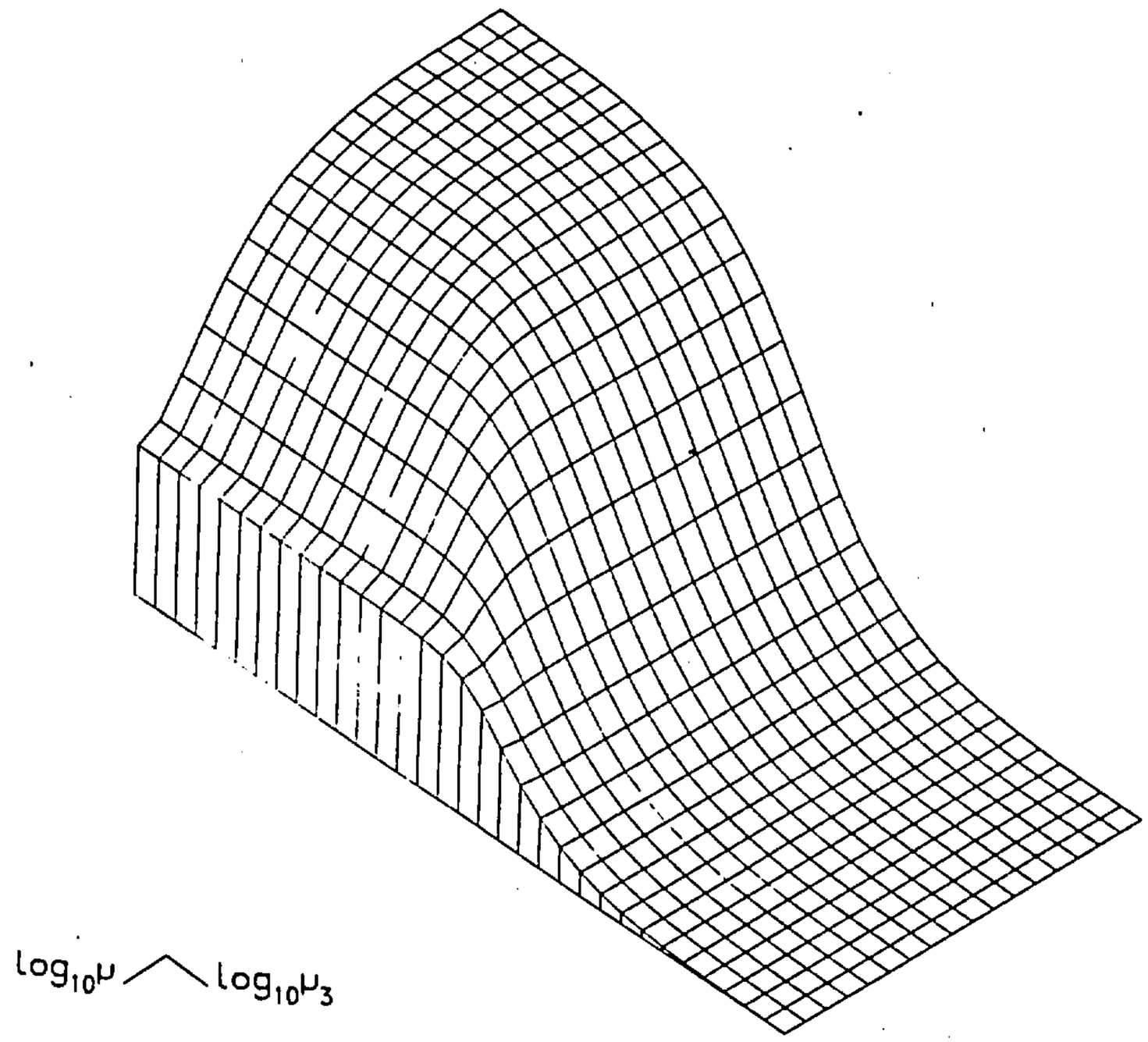


Figure 3.3: The surface of $\alpha_c = \alpha_c(\mu,\mu_3)$ for direct systems. It is described by a contour plot and a 3-dimensional plot drawn to the same logarithmic scale. $0 \lesssim \alpha_c \lesssim 1$. The marked contours, numbered 1 to 12, have the values, 0.01, 0.1, 0.2,...,0.8,0.9,0.95,0.99.

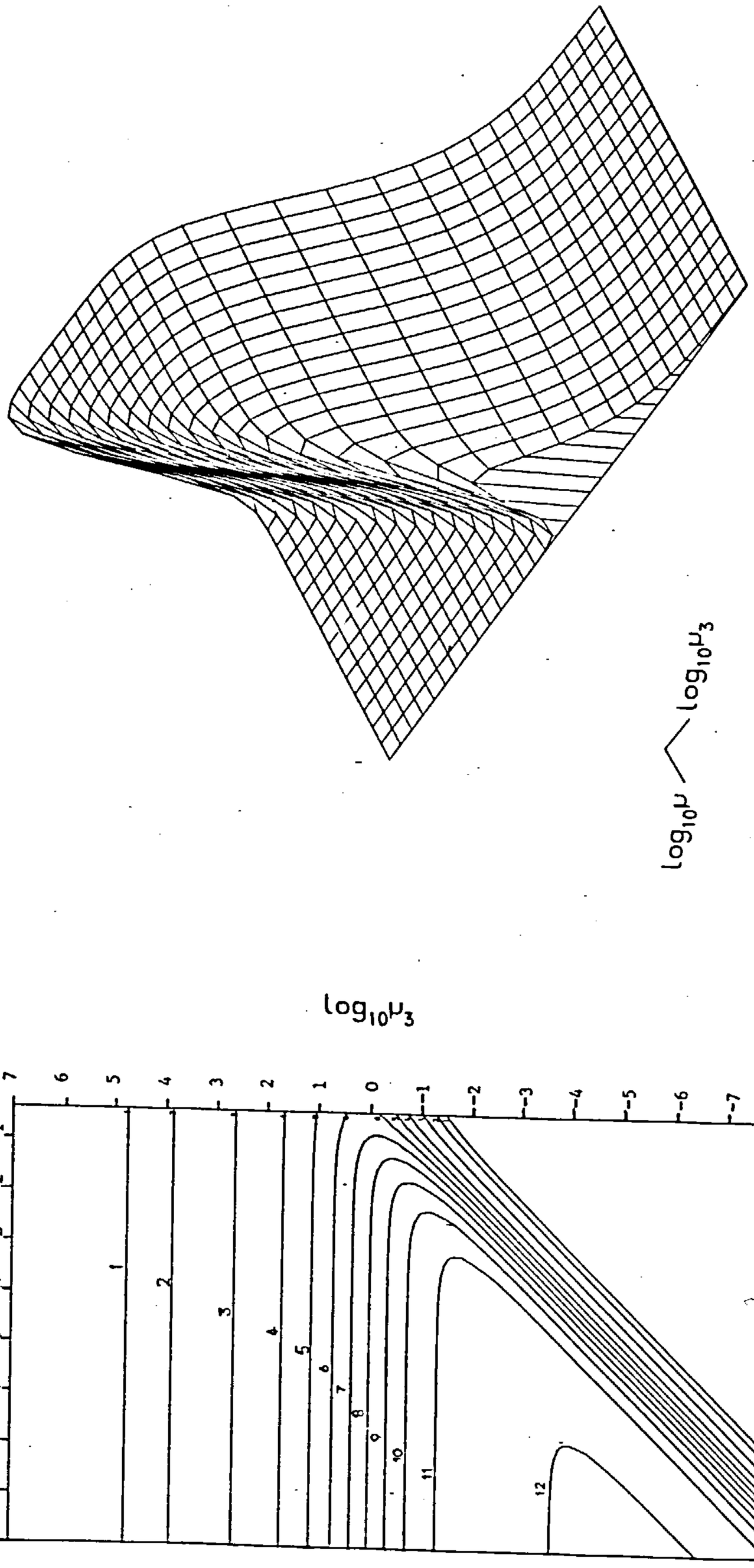


Figure 3.4: The surface of $\alpha = \alpha(\mu,\mu_3)$ for retrograde systems. $0 \leq \alpha \leq 0.25$. The marked contours, numbered 1 to 12^c, have the values, 0.005, 0.01, 0.025, 0.050^c, 0.075, ..., 0.2, 0.225, 0.249.

3.3

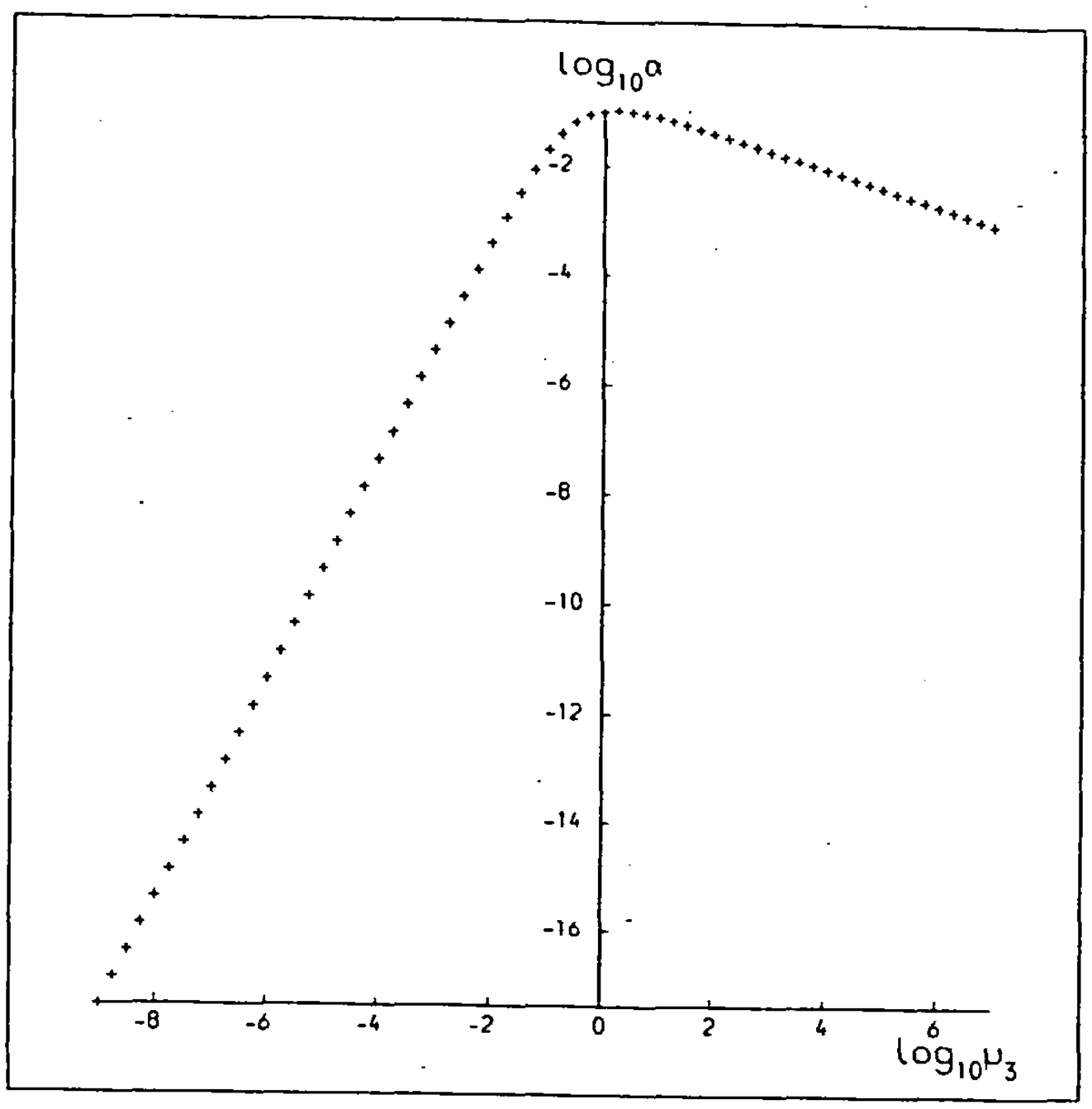


Figure 3.5: $\log \alpha_c$ against $\log \mu_3$ for $\mu = 0.5$. For $\mu_3 << 1$, $\alpha_c \sim \mu_3^2$.

that does not correspond to real systems (μ,μ_3,α) when $\alpha = \alpha_c$. The region has as a boundary, the curve $\varepsilon_{32} = \varepsilon_{32}(\varepsilon_{23},\alpha_c)$ when $\mu = \frac{1}{2}$, and μ_3 is varied.

It is seen that for the direct systems (Figure 3.6), $\varepsilon_{23} \to \mu$, $\varepsilon_{32} \to \mu_3$, as $\mu,\mu_3 \to 0$, thus preserving the rectangular look of the projected gridpoints. This is a consequence of the fact that $\alpha \to 1$ as $\mu,\mu_3 \to 0$. For the retrograde systems, $\alpha \not\to 1$ as $\mu,\mu_3 \to 0$. Hence the transformation assumes a more irregular shape.

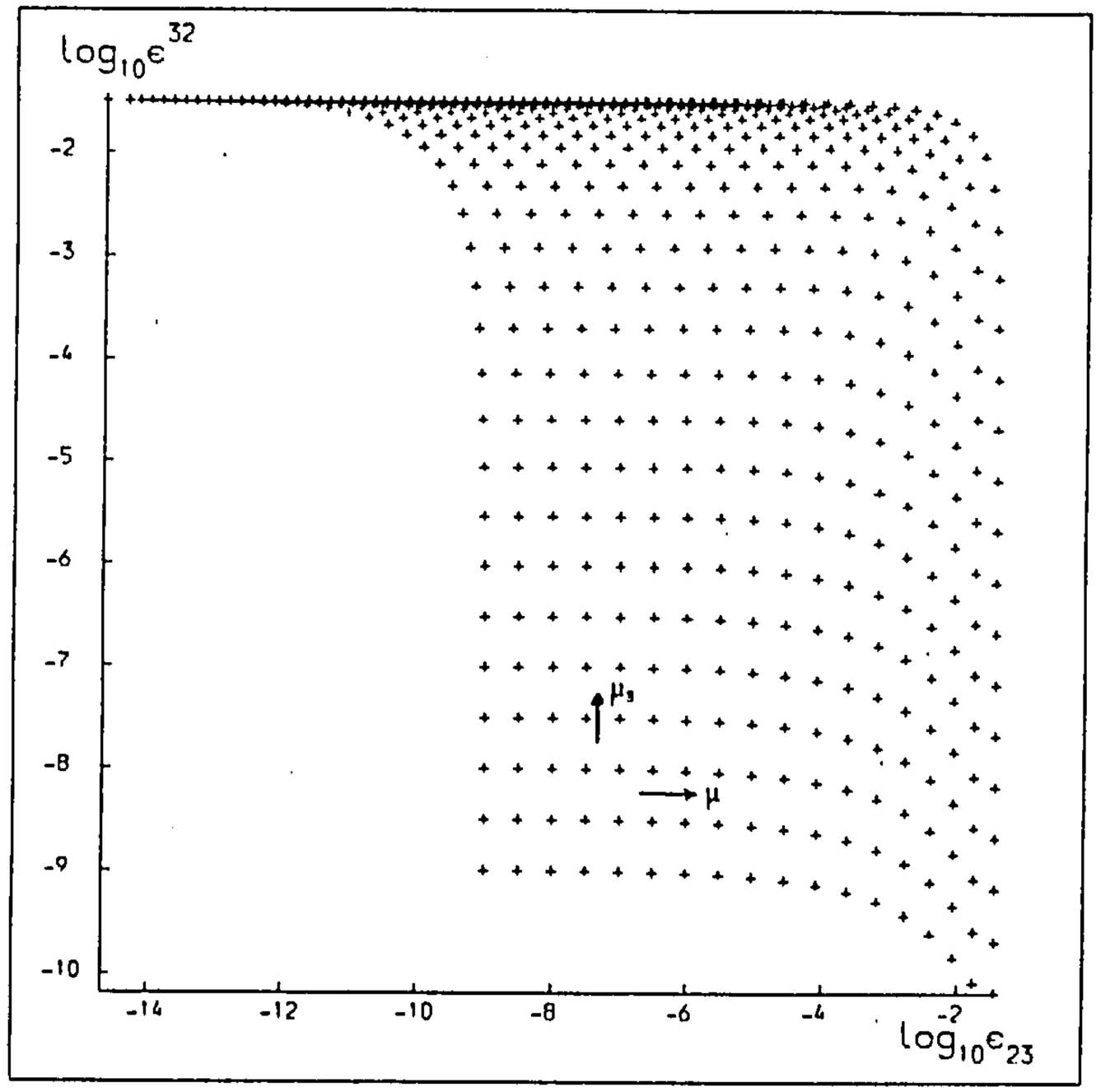


Figure 3.6: For direct systems, the transformation of a rectangular grid of (μ,μ_3) points onto the $(\varepsilon^{23},\varepsilon_{32})$ space when $\alpha = \alpha_c$. The scales are logarithmic with $\log_{10}\mu = -9, -8.5,\ldots,-0.5$ and $\log_{10}\mu_3 = -9, -8.5, \ldots, 6.5, 7$. The μ arrow indicates increasing μ with μ_3 fixed; the μ_3 arrow indicates increasing μ_3 with μ fixed.

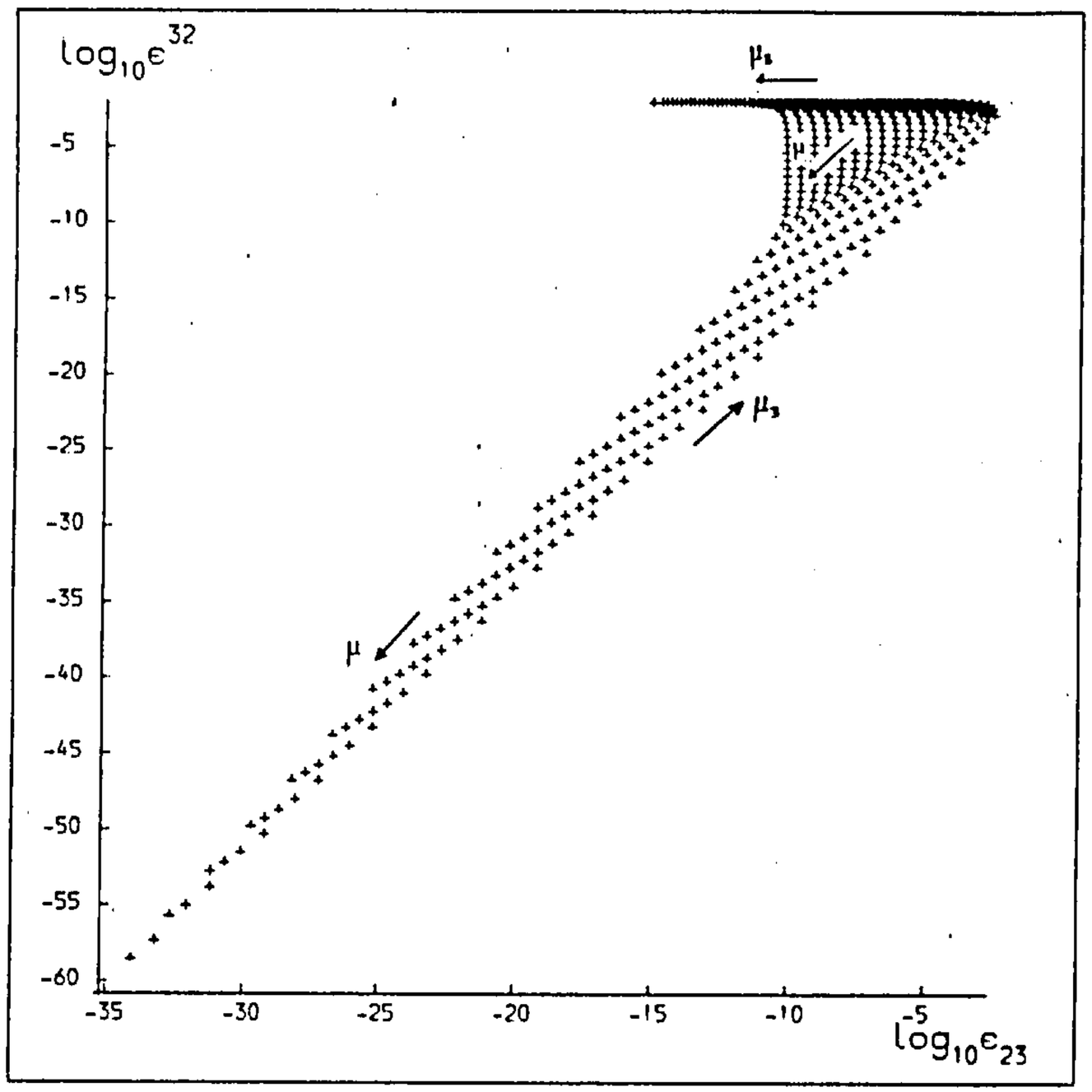


Figure 3.7: The equivalent diagram to Figure 3.6, for retrograde systems.

3.3

There are various asymptotic limits present under this transformation, the derivations of which are described in Appendix B. As $\mu_3 \to \infty$, $\varepsilon_{32} \to 3.02 \times 10^{-2}$ for direct systems and $\varepsilon_{32} \to 7.10 \times 10^{-3}$ for retrograde systems. For $\mu = 0.5$, $\mu_3 \to 0$, $\varepsilon^{23} \to 0.048$ for direct systems and $\varepsilon_{32} \sim 2^{5/2}\ \varepsilon_{23}^{7/4}$ for retrograde systems.

The contours of α_c for direct and retrograde systems in $(\varepsilon^{23}, \varepsilon_{32})$ space are given in Figures 3.8 and 3.9 respectively. Although α_c monotonically increases to unity as $\varepsilon^{23}, \varepsilon_{32} \to 0$ for direct systems this is not the case for retrograde systems.

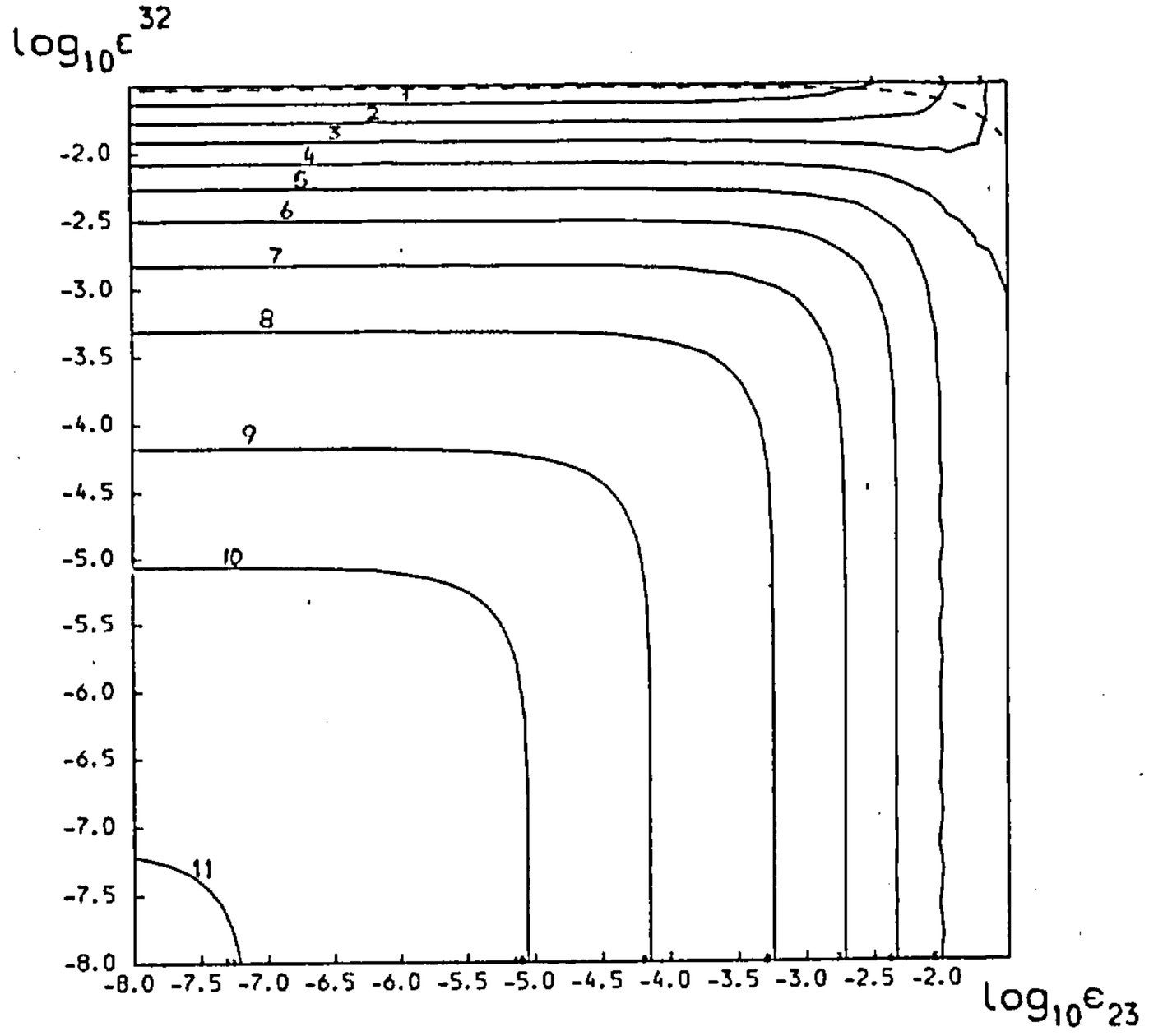
The diagrams show qualitatively the behaviour of α_c with ε^{23} and ε_{32}. Because of the interpolation routine used for finding contours of α_c, the errors in the actual values of α_c for individual $(\varepsilon^{23}, \varepsilon_{32})$ pairs may be fairly large. This is particularly true for the retrograde systems near $\mu = 0.5$ at the boundaries of possible $(\varepsilon^{23}, \varepsilon_{32})$ points.

In order to find the value of α_c at a particular pair ε^{23}, ε_{32}, a more refined technique is used. For the chosen values of $\varepsilon^{23}, \varepsilon_{32}$, the value of α_c is estimated roughly. This may be done by using the contour plots. Alternatively, if α_c is being determined for a lattice of $(\varepsilon^{23}, \varepsilon_{32})$ points, then values of α_c for neighbouring lattice points may provide a good first estimate. Having estimated α_c, μ and μ_3 may be calculated, from Equations (30) and (31), viz.

$$\mu = \frac{1}{2}\left(1 - \left[1 - \frac{4\varepsilon^{23}}{\alpha^2}\right]^{1/2}\right) \qquad (32)$$

$$\mu_3 = \varepsilon_{32}/\alpha^3 \qquad (33)$$

with $\alpha = \alpha_c$. The procedure is now the same as before. x is calculated by the Newton-Raphson method applied to Equation (28). f^2g is calculated



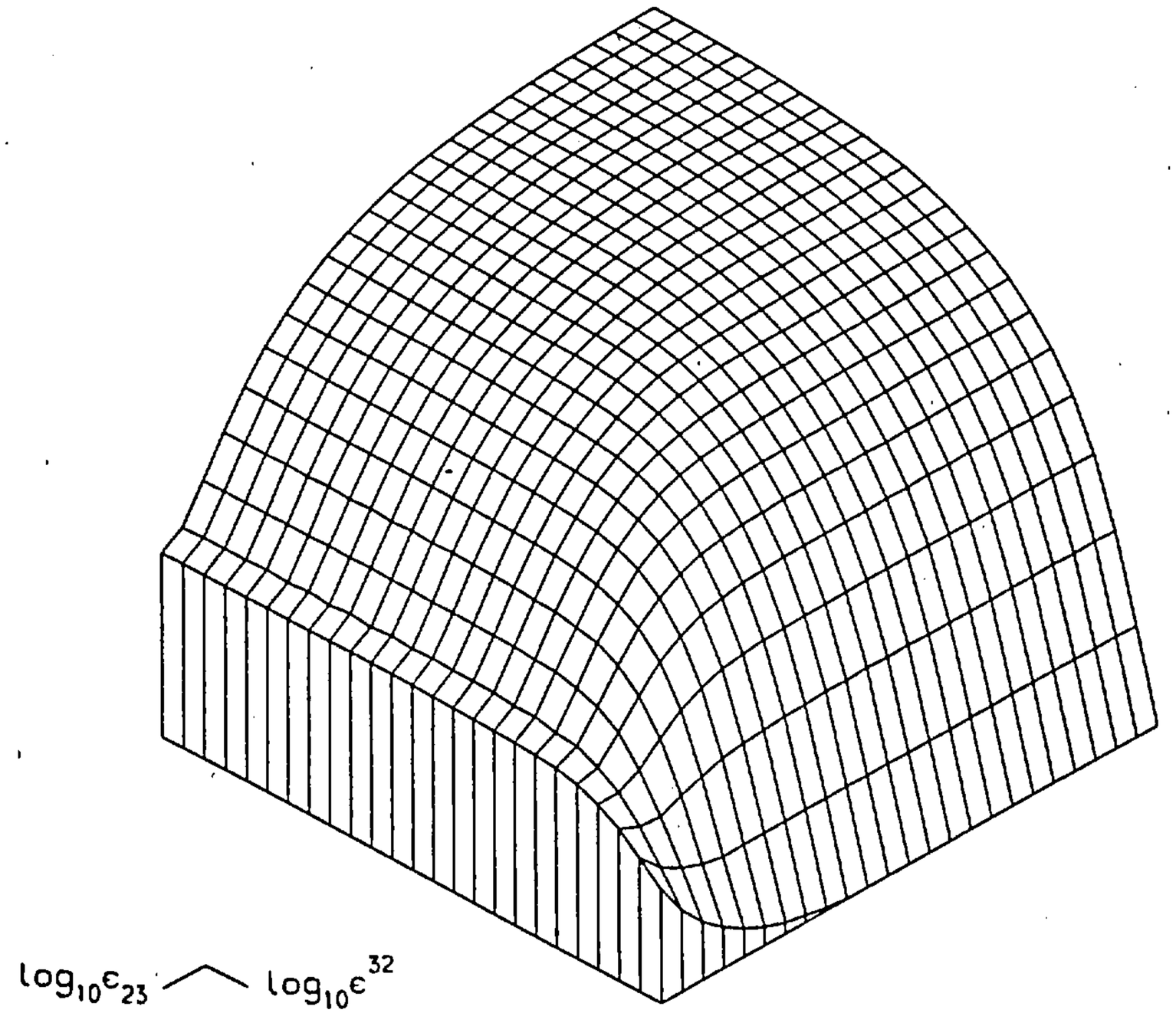


Figure 3.8: The surface of $\alpha_c = \alpha_c(\varepsilon^{23}, \varepsilon_{32})$ for direct systems. It is described by a contour plot and a 3-dimensional plot, drawn to the same logarithmic scale. $0 \lesssim \alpha_c \lesssim 1$. The marked contours, numbered 1 to 11, have the values, 0.1, 0.2,...,0.9,0.95, 0.99. The dotted line represents the values of $(\varepsilon^{23}, \varepsilon_{32})$ at $\mu = 0.5$, $\alpha = \alpha_c$.

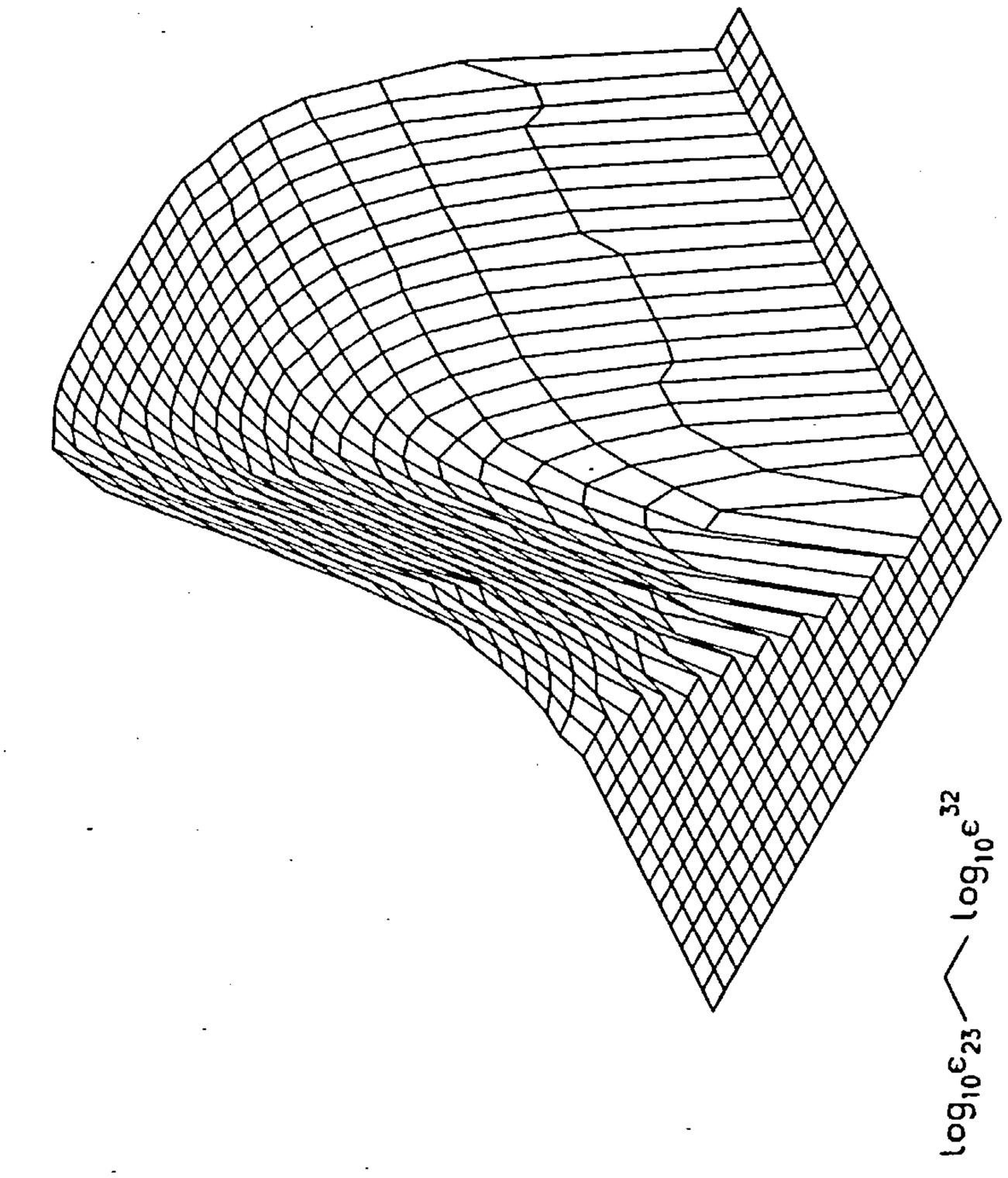


Figure 3.9: The surface of $\alpha_c = \alpha_c(\varepsilon^{23},\varepsilon^{32})$ for retrograde systems. $0 \lesssim \alpha_c \lesssim 0.25$. The contours, numbered 1 to 9, have the values 0.05, 0.075, ..., 0.2, 0.225, 0.24.

3.3

from Equations (22), (24), (25). From Equation (29), a new value of α_c is calculated by Newton-Raphson. The procedure is then repeated, by recalculating μ, μ_3 from Equations (32), (33), and so on, until the value of α_c has converged to sufficient accuracy.

This procedure is performed for a range of $\varepsilon^{23}, \varepsilon_{32}$ for direct and retrograde systems and are presented in Tables 3.1, 3.2.

ε_{32}					
10^{-2}	0.502	0.502	0.499	0.474	0.291
10^{-3}	0.759	0.758	0.750	0.681	0.343
10^{-4}	0.888	0.884	0.856	0.734	0.355
10^{-5}	0.946	0.934	0.881	0.740	0.356
10^{-6}	0.970	0.946	0.885	0.741	0.356
	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2} ε^{23}

Table 3.1: Values of α_c for various $\varepsilon^{23}, \varepsilon_{32}$, for direct systems.

ε_{32}					
10^{-2}	-	-	-	-	-
10^{-3}	0.212	0.212	0.203	0.145	-
10^{-4}	0.241	0.233	0.180	-	-
10^{-5}	0.239	0.188	0.081	-	-
10^{-6}	0.189	0.084	-	-	-
	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2} ε^{23}

Table 3.2: Values of α_c for various $\varepsilon^{23}, \varepsilon_{32}$, for retrograde systems. A dash indicates that no value of α_c exists.

3.3

The values of α_c for direct systems are slightly smaller than those calculated by Walker and Roy (Paper III), since they used the exact expression for c^2h rather than the two-body approximation of Szbehely and Zare.

3.4 Discussion

This chapter has been concerned with examining the conditions for ensuring hierarchical preservation of three-body systems. For coplanar, initially circular systems, critical values α_c have been determined such that any system with $\alpha < \alpha_c$ is guaranteed to be hierarchically preserved for all time. This does not mean that such a system is hierarchically stable. For example, the criterion does not rule out the possibility of the outer body m_3, escaping the system. Marchal (1985) shows possible limits of bounded motion, and describes separate tests for the eventual escape of the outer body.

Intuitively, it is felt that α_c should increase as ε^{23} and ε_{32} decrease. In other words, as the perturbations on the two-body systems decrease, m_2 can stand being closer to m_3 without their proximity drastically affecting each others orbit. Indeed, if there are no perturbations at all ($\varepsilon^{23} = \varepsilon_{32} = 0$), the Keplerian orbits would remain unchanged, no matter how close m_2 was to m_3, in which case α may exceed 1.

This reasoning is reflected for direct systems by $\alpha_c \to 1$ as ε^{23}, $\varepsilon_{32} \to 0$. The same is not true of retrograde systems. For constant ε^{23} , ε_{32} (or constant μ,μ_3), the retrograde value of α_c is always less

3.4

than the direct value. For many values of ϵ^{23}, ϵ_{32} there is no guarantee of hierarchical preservation no matter how small the value of α is, for the retrograde case.

Two questions arise out of this chapter. The first is whether all systems with no guarantee of hierarchical preservation are hierarchically unstable or not. The analytical criterion indicates regions in the configuration space which guarantee hierarchical preservation. Is there a larger region in the configuration space, inside which all systems are hierarchically stable? The second question is this: given that such an empirical stability region exists, is it larger for direct systems than it is for retrograde systems, or vice versa?

There is some evidence to support the view that an empirical stability region outside the analytical region does exist. The analytical criterion is based on the fact that the minimum of the Sundman function ($\sqrt{p/a}$) is greater than $\sigma/\nu(L_1)$. There is no reason to suppose that if this condition is not satisfied, that the actual value of σ/ν for a system must be less than $\sigma/\nu(L_1)$, at some time in its evolution. Even if $\sigma/\nu < \sigma/\nu(L_1)$ at some time, this does not mean that the outer body must come close to one of the other bodies. Thus the criterion, although obviously sufficient for hierarchical preservation, may not be necessary.

For the restricted problem, the Hill stability curves provide a similar criterion for stability to the c^2h criterion. However, it can be shown that there are regions where stable periodic orbits exist which are not Hill stable, (e.g. Markellos, 1973). Thus an empirical stability region is possible in the restricted problem.

Nacozy (1977) found that the Sun-Jupiter-Saturn system is stable even if the planetary masses are multiplied by a factor up to about 30. The

3.4

analytical criterion only guarantees hierarchical preservation for planetary masses up to 25 times their actual value. Walker and Roy (Paper III) indicated thatan empirical stability region exists for coplanar, initially circular, direct, three-body systems, by performing many numerical integration experiments. Thus there is some experimental evidence for the existence of an empirical stability region in the general three-body problem.

The methods used by Walker and Roy will be used in this work. Their results for direct systems will be confirmed and it will be shown that a similar empirical region of stability exists for retrograde systems.

If the surfaces of $\alpha_c = \alpha_c\ (\varepsilon^{23}, \varepsilon_{32})$ are to be taken at face value, then it seems likely that the empirical stability region for direct systems is larger than that for retrograde systems. We shall see in Chapters 4 and 6 that this is not the case. In fact, for constant ε^{23}, ε_{32}, α, the retrograde system is more stable than its direct counterpart.

CHAPTER 4

NUMERICAL EXPERIMENTS FOR FICTITIOUS RETROGRADE THREE-BODY SYSTEMS

4.1 Introduction

In the previous chapter, sufficient conditions were described for a three-body system to have its given hierarchy preserved for all time. These conditions corresponded to regions in the configuration space. The values of α_c obtained reflect qualitatively the expected behaviour of direct systems, i.e. α_c increases as ε^{23} ,ε_{32} decrease. This is not the case for the retrograde systems. Thus either our intuitive ideas concerning the stability of retrograde systems is wrong, or the analytical stability criterion is not an adequate tool for our examination.

This is allied with the fact that the definition of hierarchical stability given in Chapter 2, is a more stringent concept than the hierarchical preservation of Chapter 3. In the absence of any rigorous analytical treatment of hierarchical stability, the easiest approach is by numerically modelling individual systems.

In this chapter, particular attention is paid to fictitious retrograde systems, while direct systems are considered in Chapter 5. The numerical procedures and subsequent curve fitting techniques are based on those used by Walker and Roy, Paper III, for direct systems. These results are briefly reviewed in Section 4.2. The curve fitting techniques and their subsequent augmentation for both direct and retrograde systems are explained in Section 4.4. The results from several hundred numerical models are also presented in this section. A brief discussion of the behaviour of the systems as they approach instability is given in Section 4.5.

A historical note is in order at this point. The work on direct systems was started before the work on retrograde systems. The

4.1

reason for discussing the retrograde systems first is that in many ways the data from these systems is easier to handle and explain. Additional curve-fitting techniques are needed for the direct systems in order to predict the stable lifetimes of direct systems. This is necessary due to the effect of sommensurabilities. These added complications will be explained in Chapter 5.

4.2 The Empirical Stability Region for Direct Three-Body Systems.

In Walker and Roy, Paper III, the authors describe the results of several hundred numerical experiments for direct coplanar three-body systems. They concentrated their examination on the following values of ε^{23}, ε_{32};

$$\varepsilon^{23} = 10^{-j} \qquad j = 2,3,4,5,6.$$

$$\varepsilon_{32} = 10^{-k} \qquad k = 2,3,4,5,6.$$

for orbits that are initially circular and starting from conjunction with m_2 between m_1 and m_3. For a given pair of epsilons, the initial value of α was varied and the equations of motion for the given system were numerically integrated. The system was thus studied until it exhibited an instability according to the definition of hierarchical stability, given in Chapter 2. The lifetime of the system until instability was noted. In this way, graphs of the lifetime against α were obtained for fixed ε^{23} , ε_{32}.

The unit of time was the synodic period of the system. It was considered that the system was most highly perturbed near a conjunction, hence it seemed important to monitor the number of conjunctions that a system "survives" for. This will be discussed again in Section 4.5 for retrograde systems.

4.2

From their results, Walker and Roy concluded that for given ε^{23}, ε_{32}, the α-line could be divided into four intervals. The first interval is $[0, \alpha_c]$, where α_c is the critical value of α for a guarantee of hierarchical preservation. No unstable systems could be found in this region, thus systems that were hierarchically preserved were also seen to be hierarchically stable.

The second interval is $(\alpha_c, \alpha_o]$, where no unstable systems could be found. This is the empirical stability region that was postulated. In every case $\alpha_o > \alpha_c$ and in the case of the higher epsilon values, there is a considerable difference between the two values.

The third interval is (α_o, α_x) where α_x is the value of α such that m_3 may be closer to m_1 than m_2, rendering the given hierarchy as immediately broken. Within this interval, one can observe unstable systems. The criteria for instability are explained in more detail in the next section. Within this interval, the lifetimes of the systems tend to infinity as α decreases to α_o, although not monotonically so, due to the presence of commensurabilities (see Chapter 5). The trend was modelled by the function

$$N_s(\alpha) = \exp\left[\beta \left(\frac{1 - \alpha}{\alpha - \alpha_o} \right)^{\gamma} \right] - 1 \qquad \alpha_o < \alpha < 1.$$

where N_s is the stability lifetime in units of the synodic period. α_o, β, γ varied with ε^{23}, ε_{32}. This curve-fitting technique is reviewed in Section 4.4.

For $\alpha \geq \alpha_x$, the given hierarchy is meaningless, as the crossover of orbits has already taken place. The stability lifetime is zero by definition. $\alpha_x = 1/(1-\mu)$, which for small μ is only slightly

4.2

greater than one. Thus $N_s(\alpha=1)$ was assumed to be zero. It is to be expected that the results for retrograde systems which are described in this chapter will follow a similar pattern.

4.3 The Numerical Integration Routine

Several hundred different direct and retrograde three-body systems have been studied by numerical integration. All the experiments were carried out on the ICL 2988 mainframe computer at Glasgow University, using the same routine that Walker used for his experiments. In this routine, the mutual radius vectors are calculated by a tenth order Taylor series, where the derivatives are calculated by recurrence relations (Schwarz and Walker, 1982). The programme incorporates an automatic step-length regulator which shortens or lengthens the integration step of the computer in order that the error caused by truncating the Taylor series after the tenth order is less than a given tolerance, (eg. 1 part in 10^{10}).

The accuracy of the integration routine is affected by the accumulated effect of this truncation error, as well as the error caused by the computer having to round numbers to 16 significant figures. Brouwer (1937) showed that the effect of round-off error on a numerical integration after n steps is to introduce a mean error proportional to $n^{3/2}$ in the mean longitude, and to $n^{\frac{1}{2}}$ in the other orbital elements, where the error is measured in units of the last decimal place. Unfortunately this is not the whole story, since Brouwer assumes that the rounding is performed up or down in an unbiased fashion. This is not always true of modern computers, and in any case, the error is modified by the truncation error, making a rigorous analysis impossible.

4.3

We are concerned with knowing for how long we may run the numerical integration before the accumulated errors render the results as meaningless. A rough idea was gained by running the programme for fictitious three-body systems with $\epsilon^{23} = \epsilon_{32} = 0$ and initially non-circular orbits. In these cases, the elements of the two binary orbits should remain constant. Any deviation is due to the accumulated error in the routine. It was found that after 150,000 steps, the error in the position of the body in its orbit was approximately 1%. A 1% error was considered sufficiently inaccurate to stop the integration. No systems are examined beyond the 150,000 step timescale, which corresponds to lifetimes of between 4000 and 6000 synodic periods for most systems encountered.

The routine incorporates a number of diagnostics for detecting instabilities. Recalling the various criteria for hierarchical stability, (Section 2.2), the routine tests the energies of each binary orbit. If an exchange of energy between the binary orbits, results in one binary having a positive energy, then the smaller mass must be escaping the system, hence there is an instability. The routine also detects if the hierarchy is broken by a crossover of orbits, and stops the programme accordingly.

The third criterion, namely detecting an irreversible change in the size, shape or orientation of any orbit, is too difficult for the numerical procedure to handle with 100% efficiency. The best that can be done is for the routine to note wherever the running value of the eccentricity exceeds the previous maximum by 20%. The maximum is then readjusted to its new value and the routine proceeds. It is then a matter for the user to examine the output and

4.3

and decide if an increase in the eccentricity is truly irreversible. In truth this is impossible to say with any certainty. Since the eccentricities start at zero, it seem unlikely that there will ever be an irreversible decrease. Indeed this has never been experienced in any of the experiments. During the running of the programme, the eccentricities are seen to vary in a periodic manner. There may be short period fluctuations superimposed on longer period effects, but once the eccentricity has reached such a local maximum, any further increase must be viewed as a possible instability, caused for example by m_3 coming close to m_2. Such changes often precede either an escape of one of the bodies or a crossover, in which case the stability lifetime is taken up to the change in eccentricities. In other isolated cases of eccentricity changes, care must be taken to ensure that this is not part of a very long term periodic effect.

Most irreversible changes can be seen as a burst of 20% eccentricity increases taking place within a few synodic periods after a much longer period of stability. Any single occurrence of a 20% increase is normally viewed with some scepticism.

4.4 Curve-Fitting Techniques Applied to Retrograde Systems.

Over 800 fictitious retrograde three-body systems have been examined using the numerical integration procedure, described in Section 4.3. All the systems considered were coplanar, initially circular with the bodies arranged at m_1 - m_2 - m_3 conjunction at the start of the integration. Each system can be unambiguously described by m_1, m_2, m_3, a_2, a_3. By normalising the masses with respect to $m_1 + m_2$ and the distances with respect to a_3, the number of parameters

4.4

describing a system may be reduced to three, such as $\varepsilon^{23}, \varepsilon_{32}$, α . The unit of time is taken to be the synodic period of the system for reasons described in Section 4.2.

Each system was numerically integrated for up to 600 synodic periods. During that time the system was checked to see if it had exhibited any instability. If this was so, then the integration was stopped and the time noted. This time (measured in synodic periods) is called the stability lifetime for that given system.

The fictitious systems that were chosen all had values of ε^{23} , ε_{32} given by

$$\varepsilon^{23} = 10^{-k} \qquad k = 2,3,4,5,6$$

$$\varepsilon_{32} = 10^{-m} \qquad m = 1,2,3,4,5,6.$$

This range reflects the range of epsilon values found in the Solar System for the planets and their natural satellites (see Chapter 6). Graphs of stability lifetime N_s against α for given $(\varepsilon^{23}, \varepsilon_{32})$ - pairs are presented in Figures 4.1. The crosses indicate the actual stability lifetime for that system and a line indicates the best fit curve through these points. This is discussed below. The circles represent systems for which no instabilities were detected during the 600 synodic period investigation time. No definitive stability lifetime can be attached to these systems. They may exhibit instabilities if we study them for long enough. Alternatively they may always be hierarchically stable. We just don't know.

There are several points that can be made about the behaviour and trends in these graphs, before discussing the curve-fitting procedure. In general the eccentricities of the inner and outer binaries

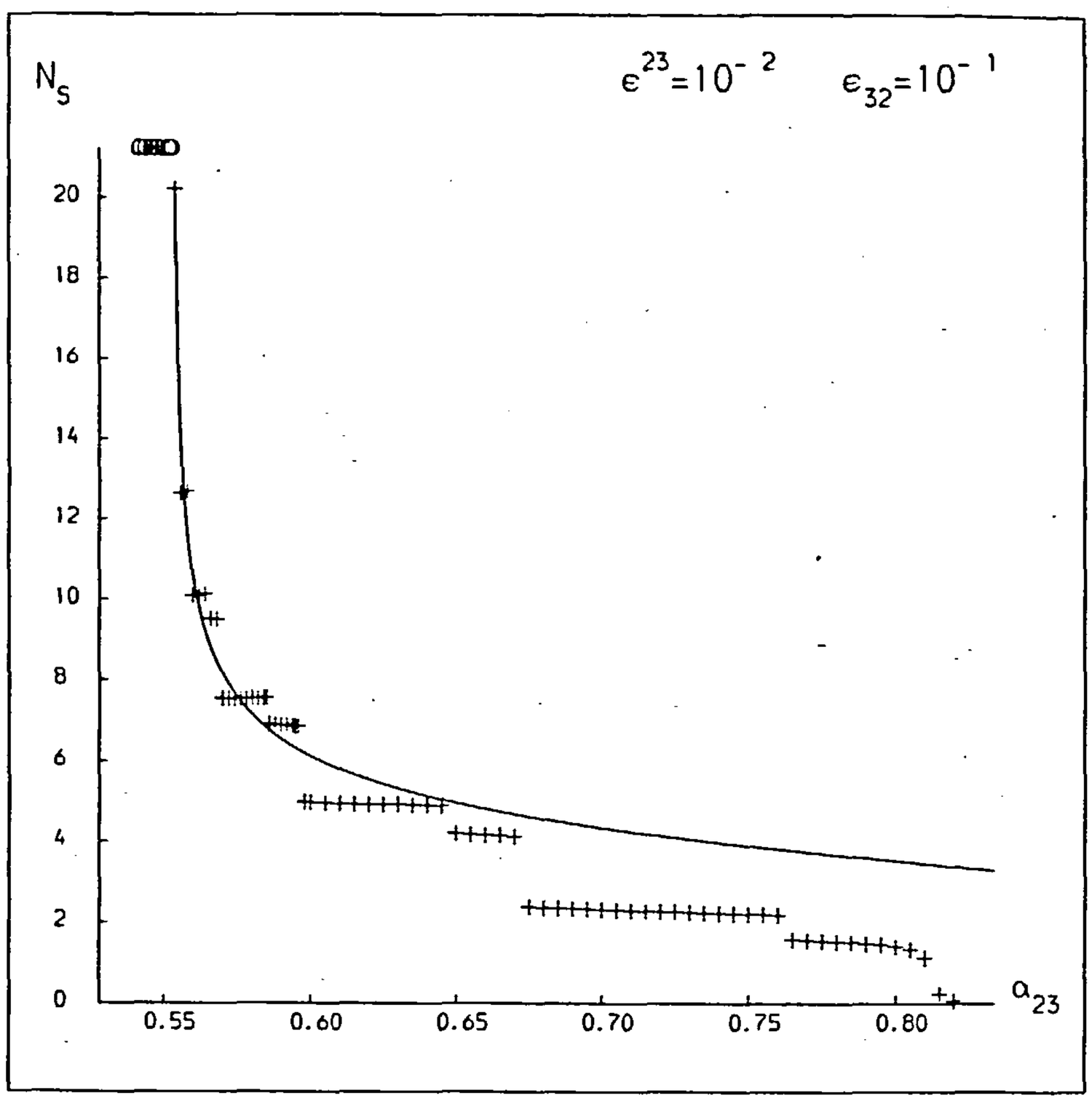
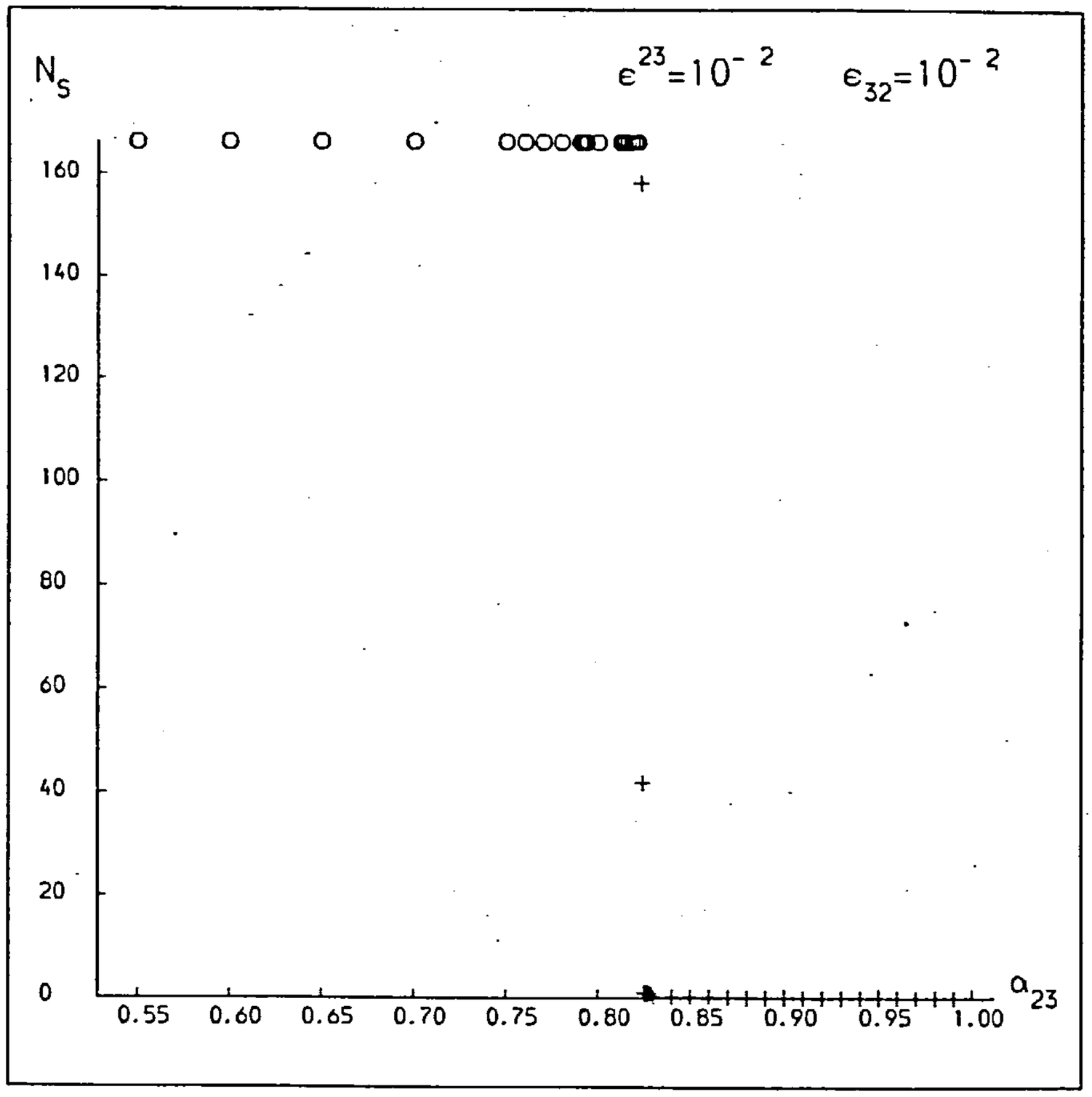
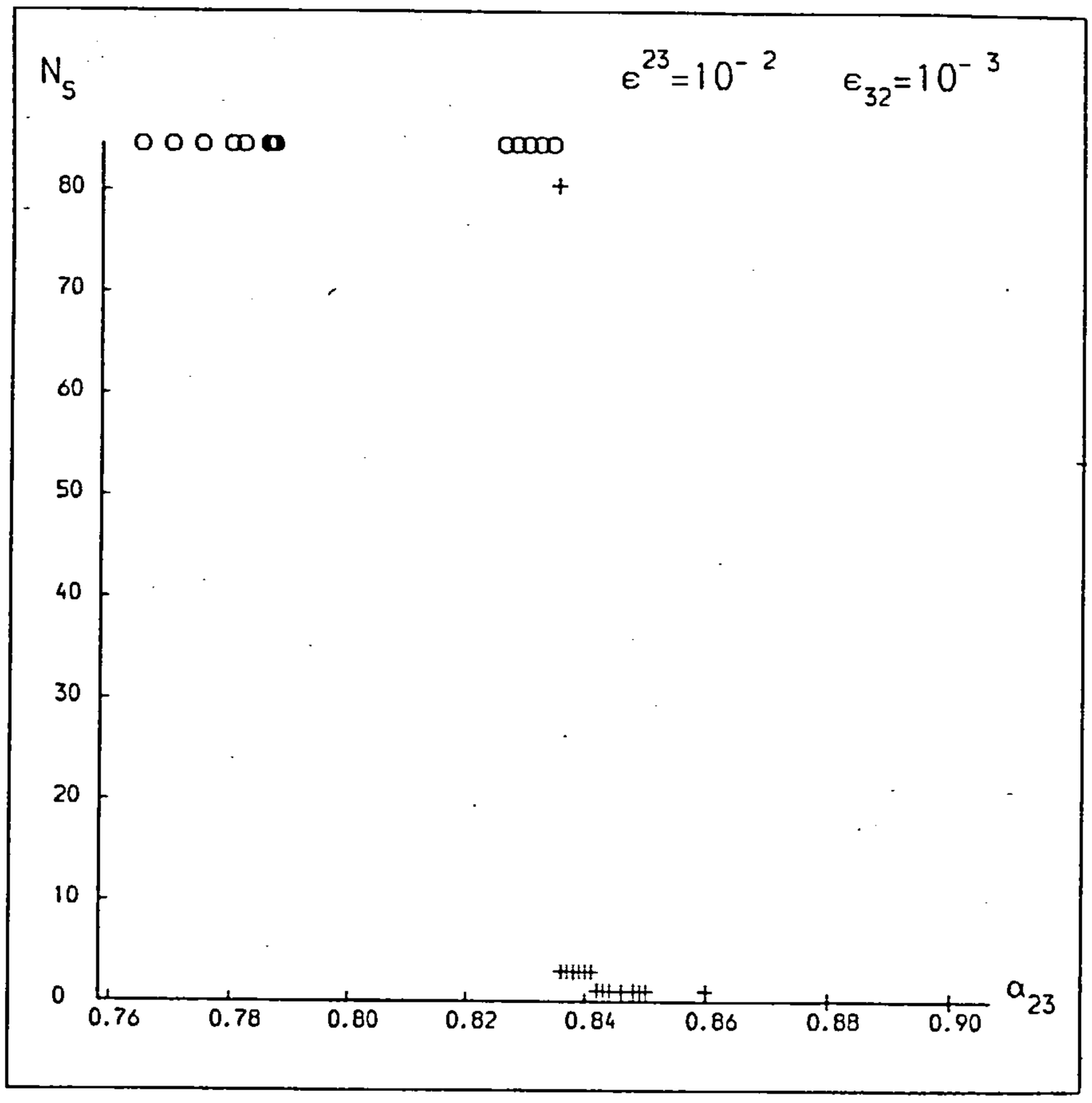
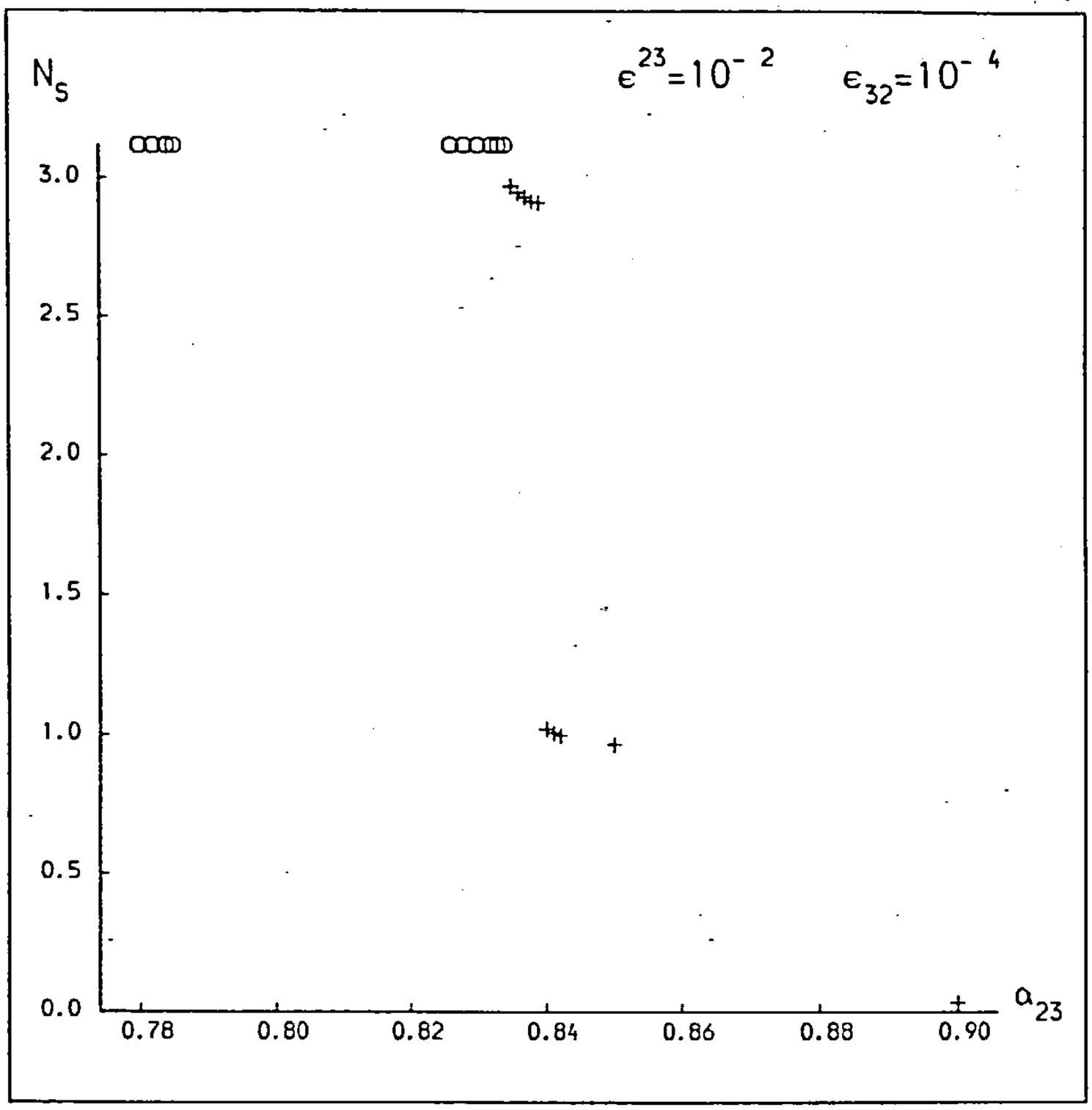


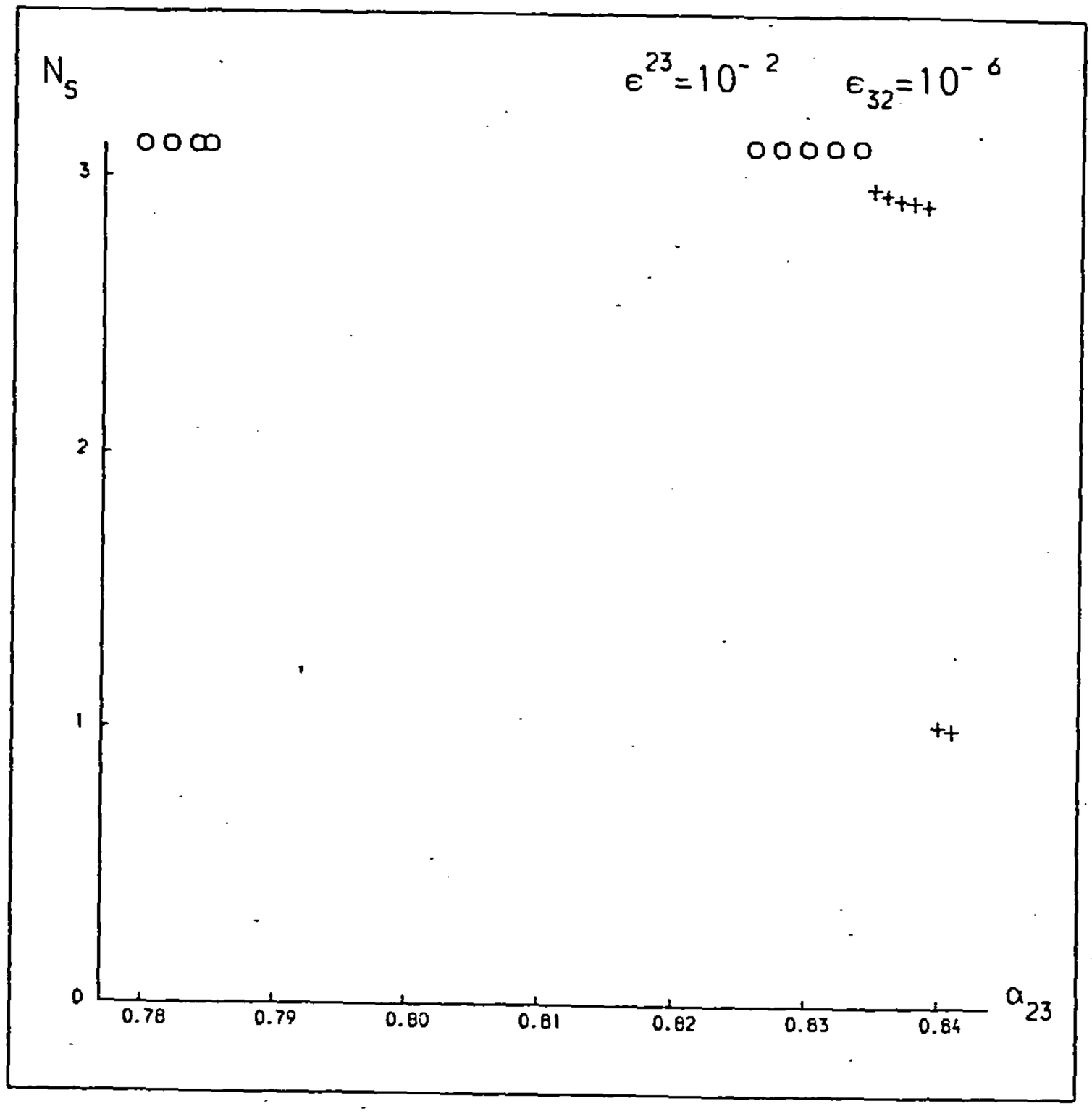
Figure 4.1: Stability lifetime against α for given ϵ^{23}, ϵ_{32}. The crosses indicate the stability lifetime for each unstable system. The circles indicate systems that have been studied for 600 synodic periods without exhibiting any instability. A line indicates a best-fit curve through the unstable system points. A full line indicates a weighted fit while a dotted line indicates an unweighted fit.

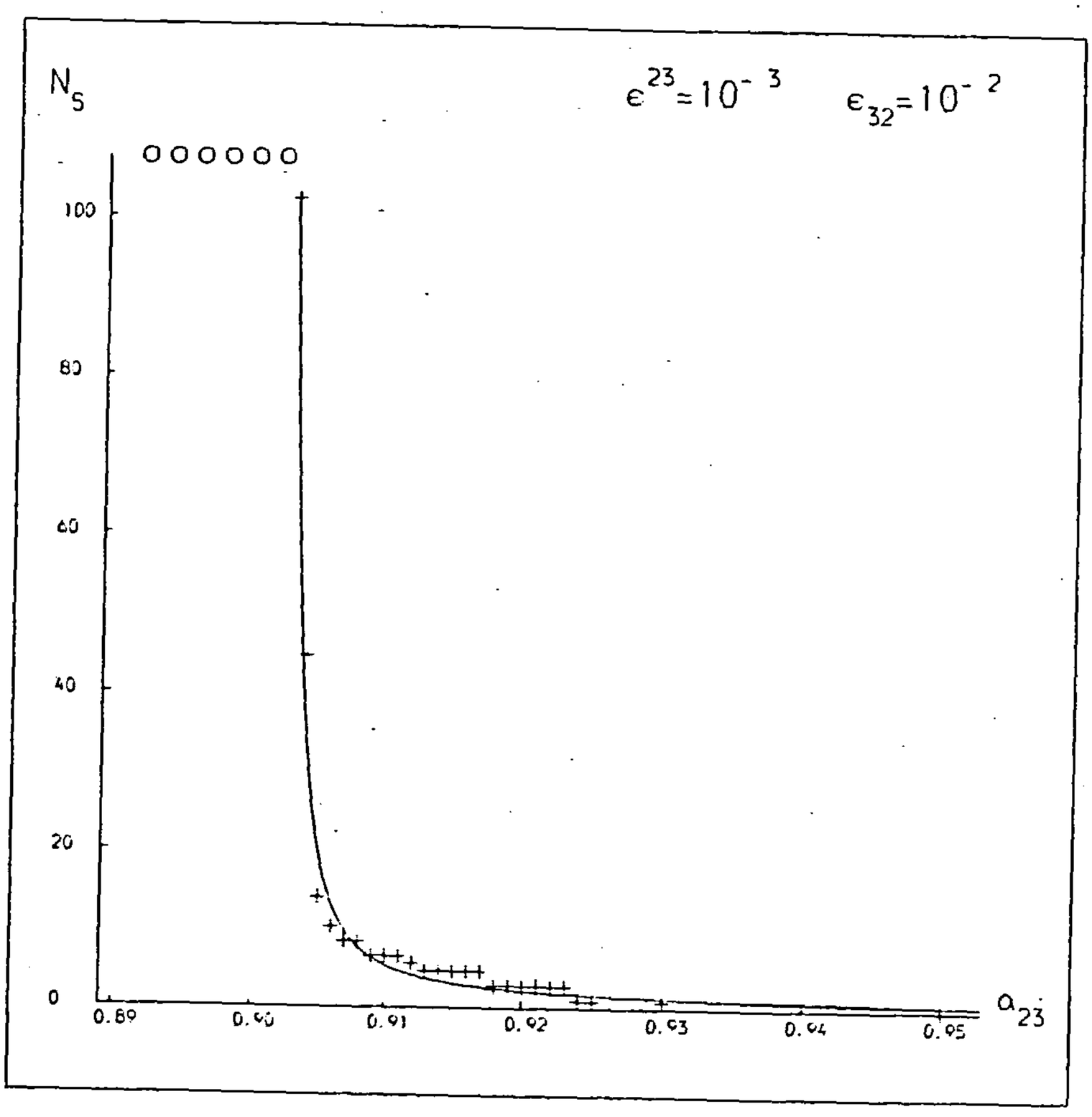

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-2}$
160
140
120
100
80
60
40
20
0
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
a_{23}

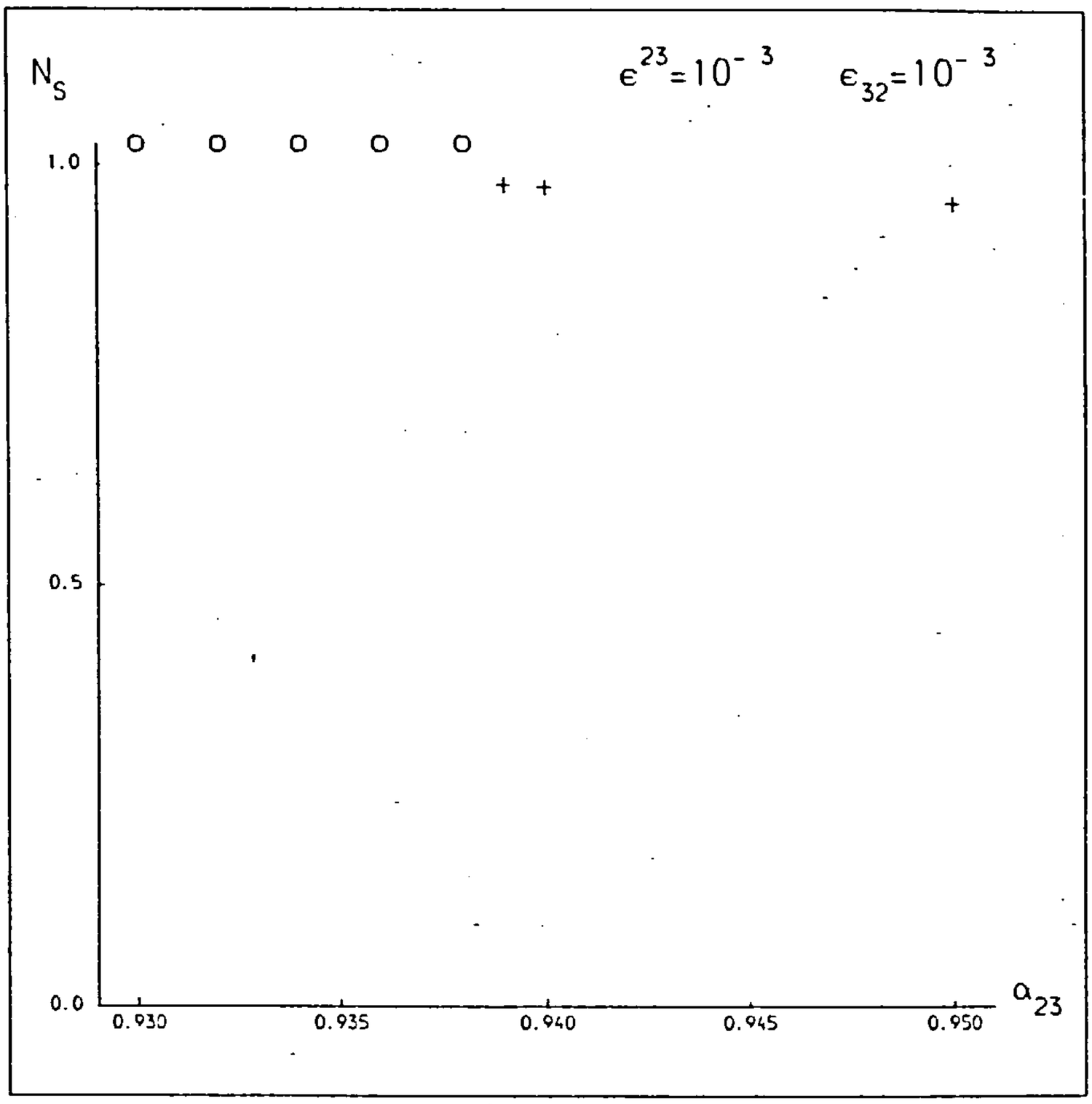

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-3}$
80
70
60
50
40
30
20
10
0
0.76
0.78
0.80
0.82
0.84
0.86
0.88
0.90
a_{23}

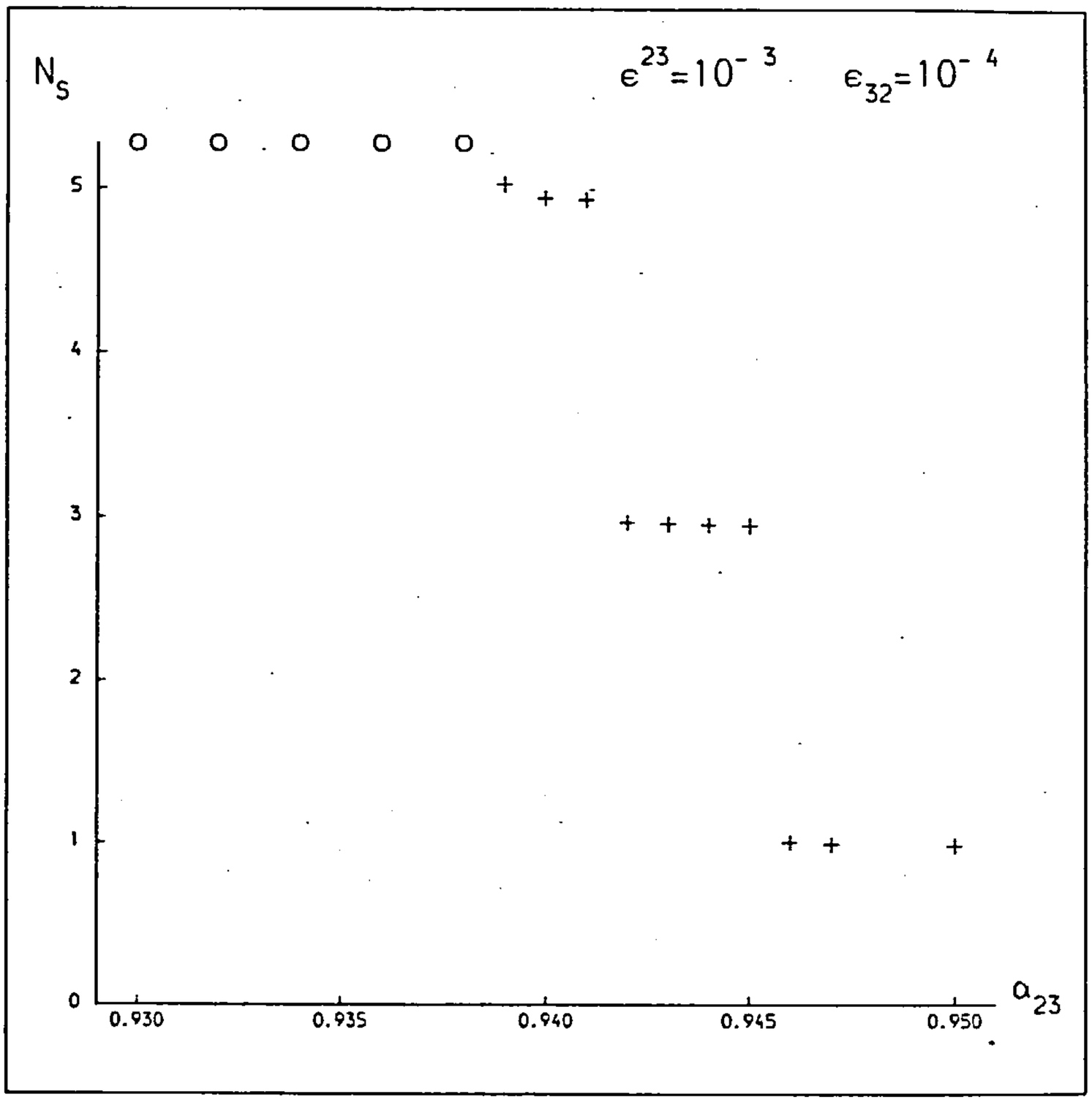

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-4}$
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.78
0.80
0.82
0.84
0.86
0.88
0.90
a_{23}

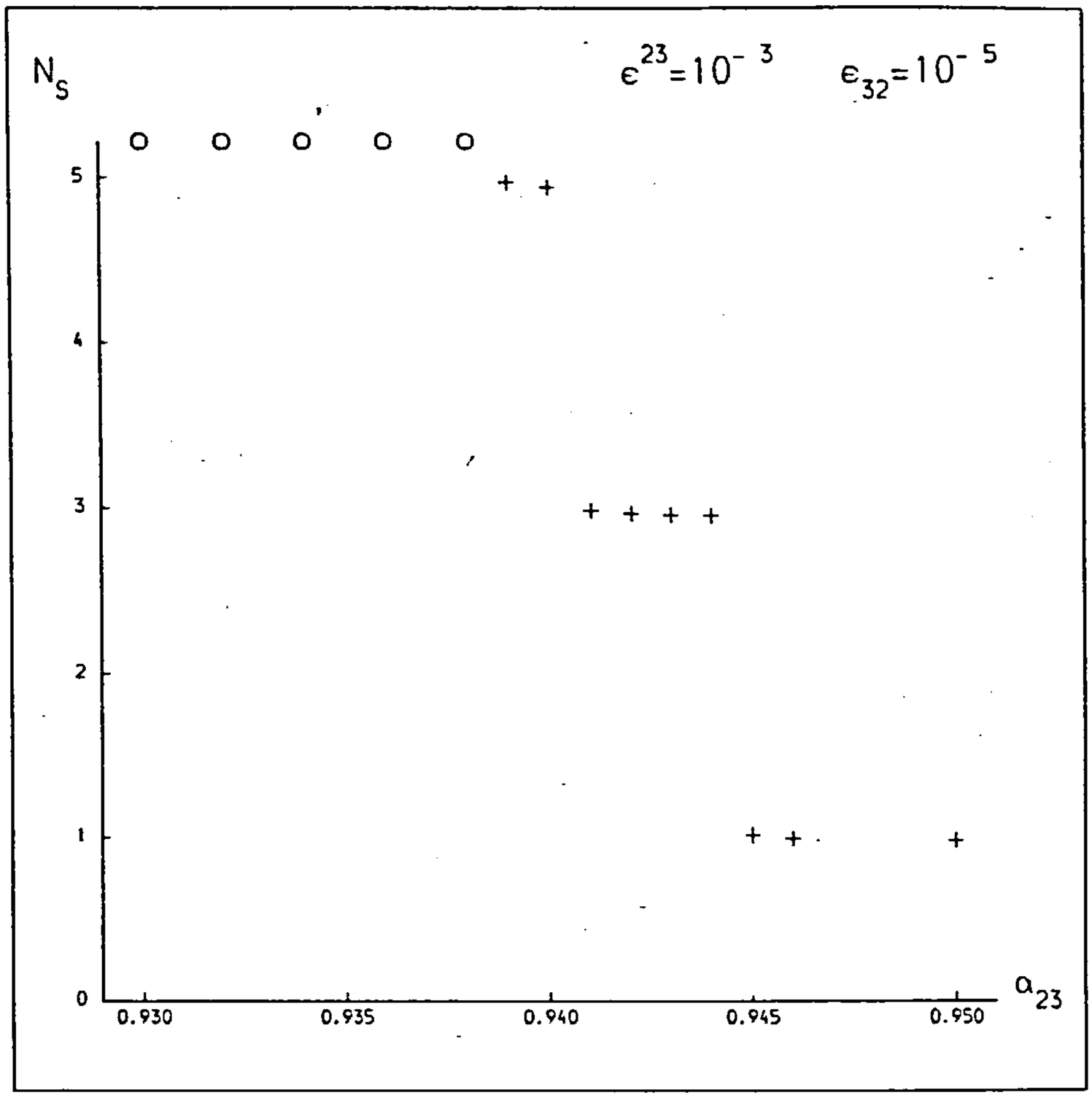
N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-5}$
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.78
0.80
0.82
0.84
0.86
0.88
0.90
a_{23}

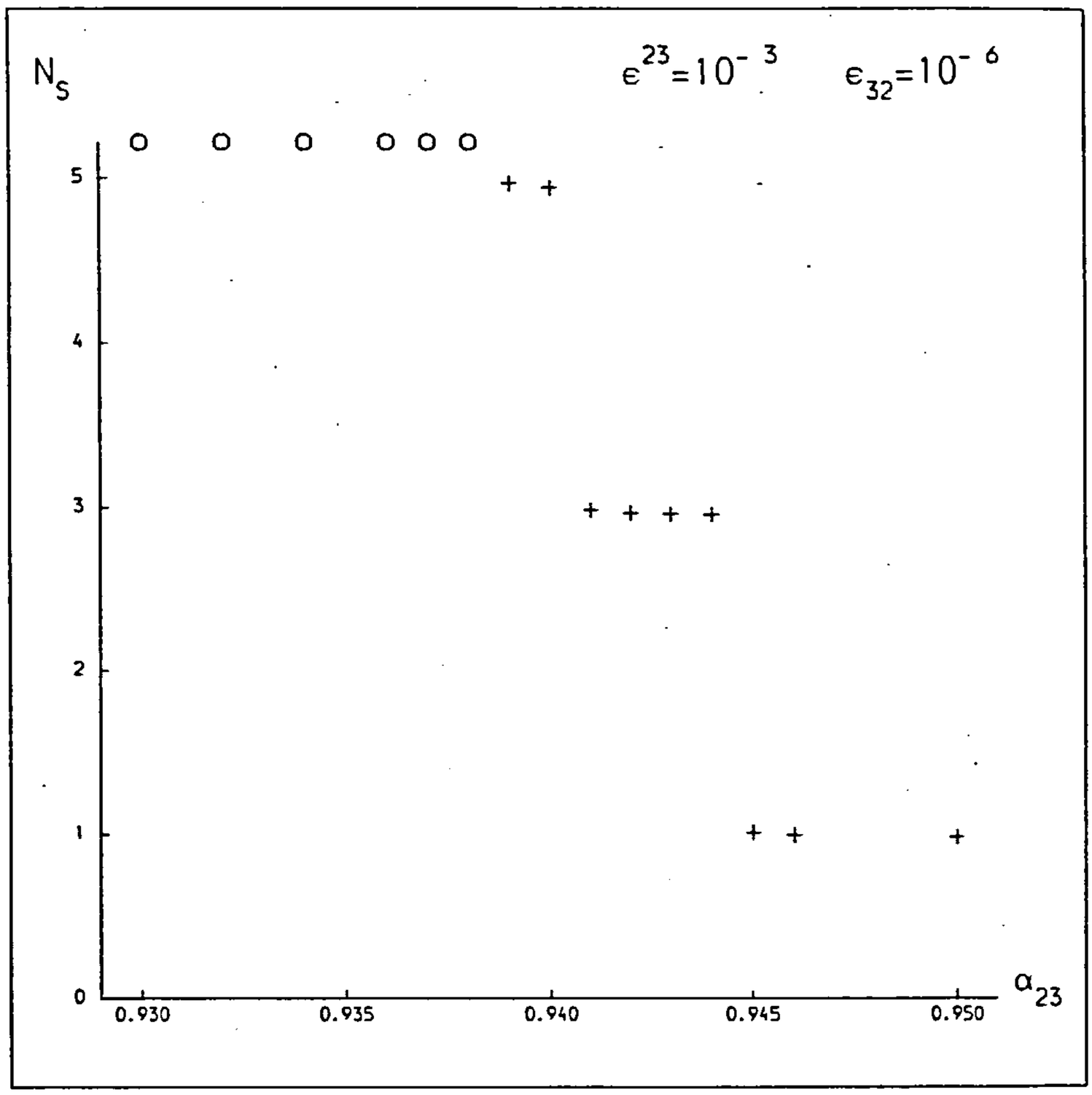

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-6}$
3
2
1
0
0.78
0.79
0.80
0.81
0.82
0.83
0.84
α_{23}

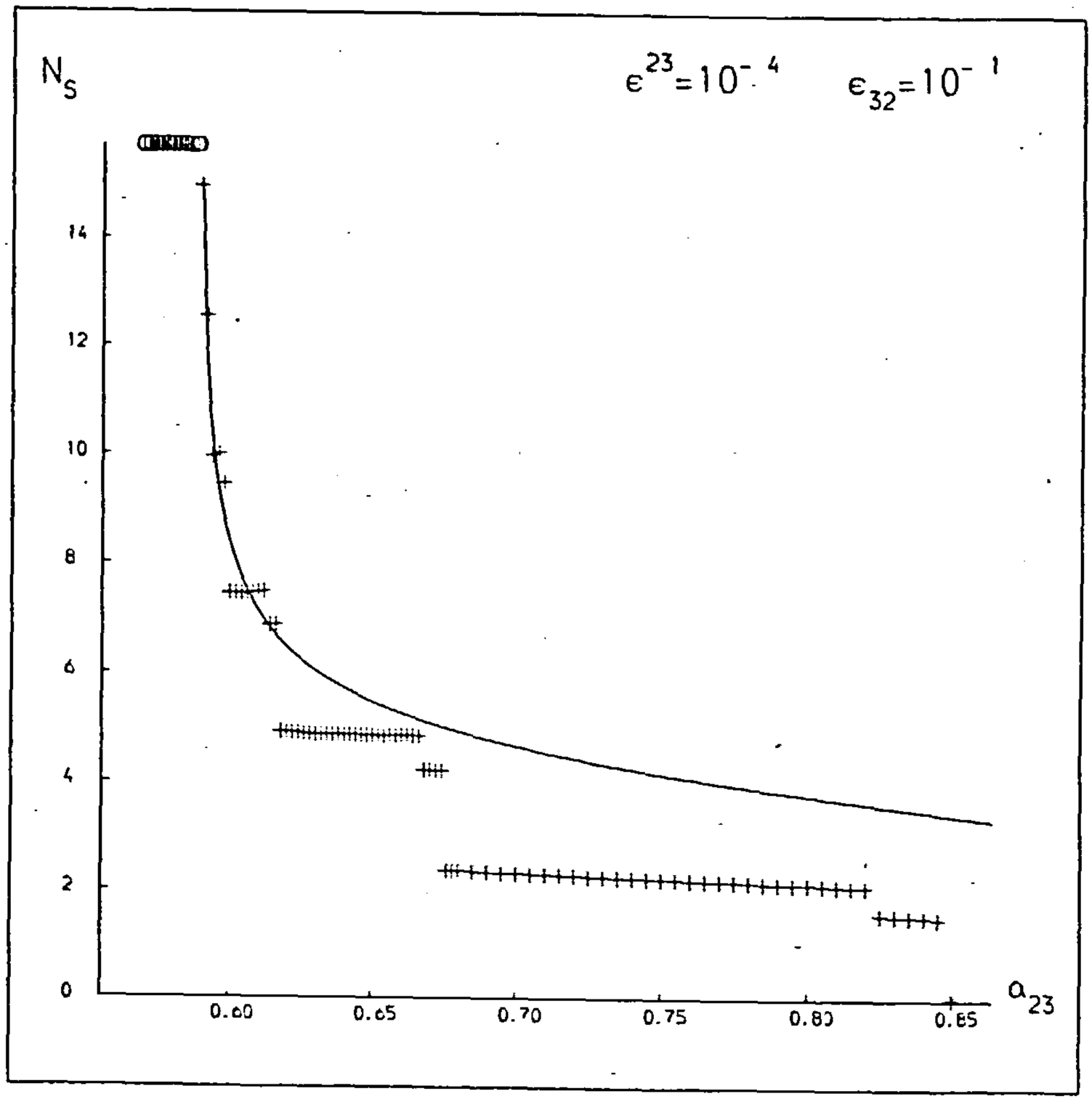

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-2}$
100
80
60
40
20
0
0.89
0.90
0.91
0.92
0.93
0.94
0.95
α_{23}

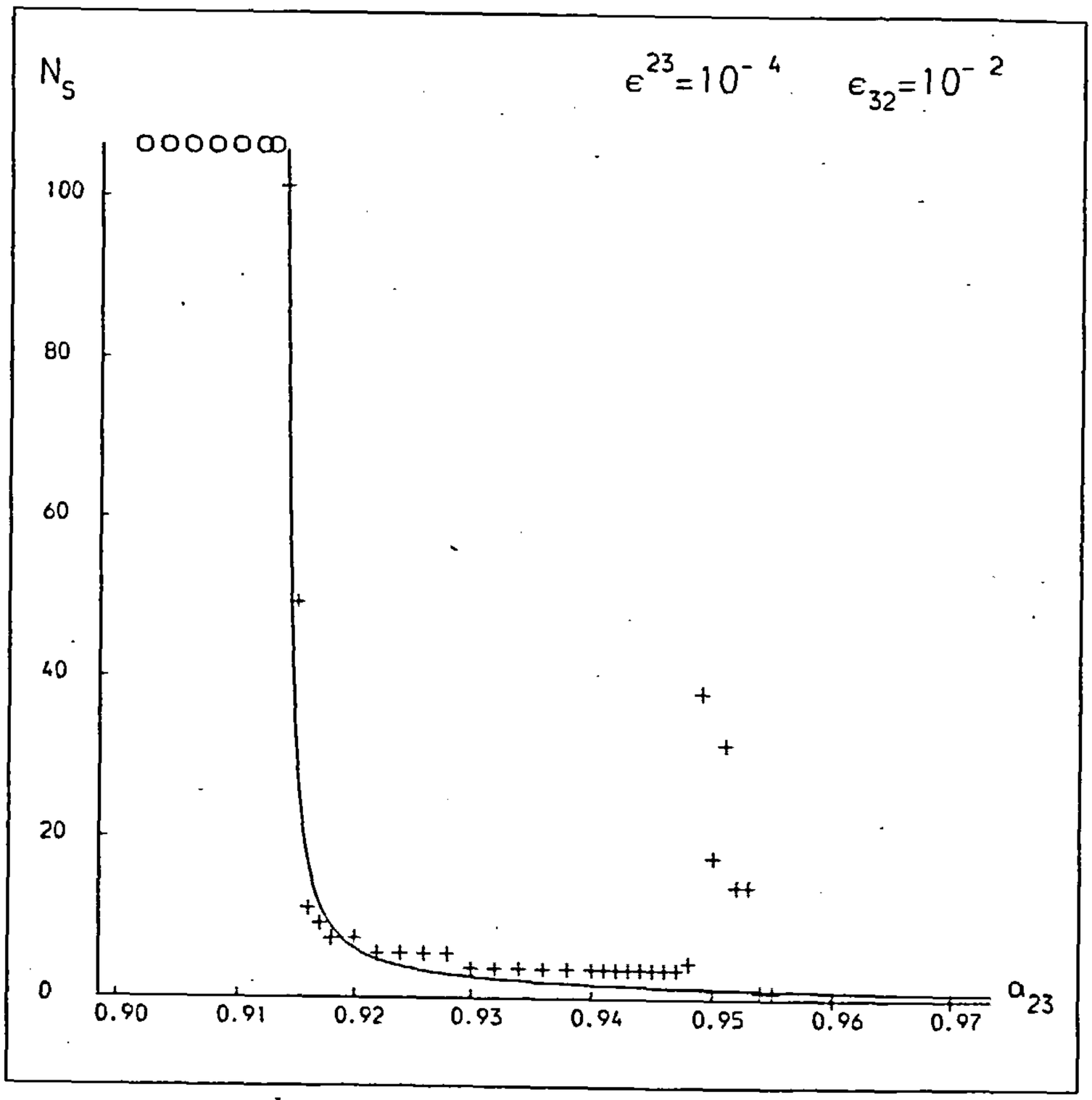

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-3}$
1.0
0.5
0.0
0.930
0.935
0.940
0.945
0.950
a_{23}

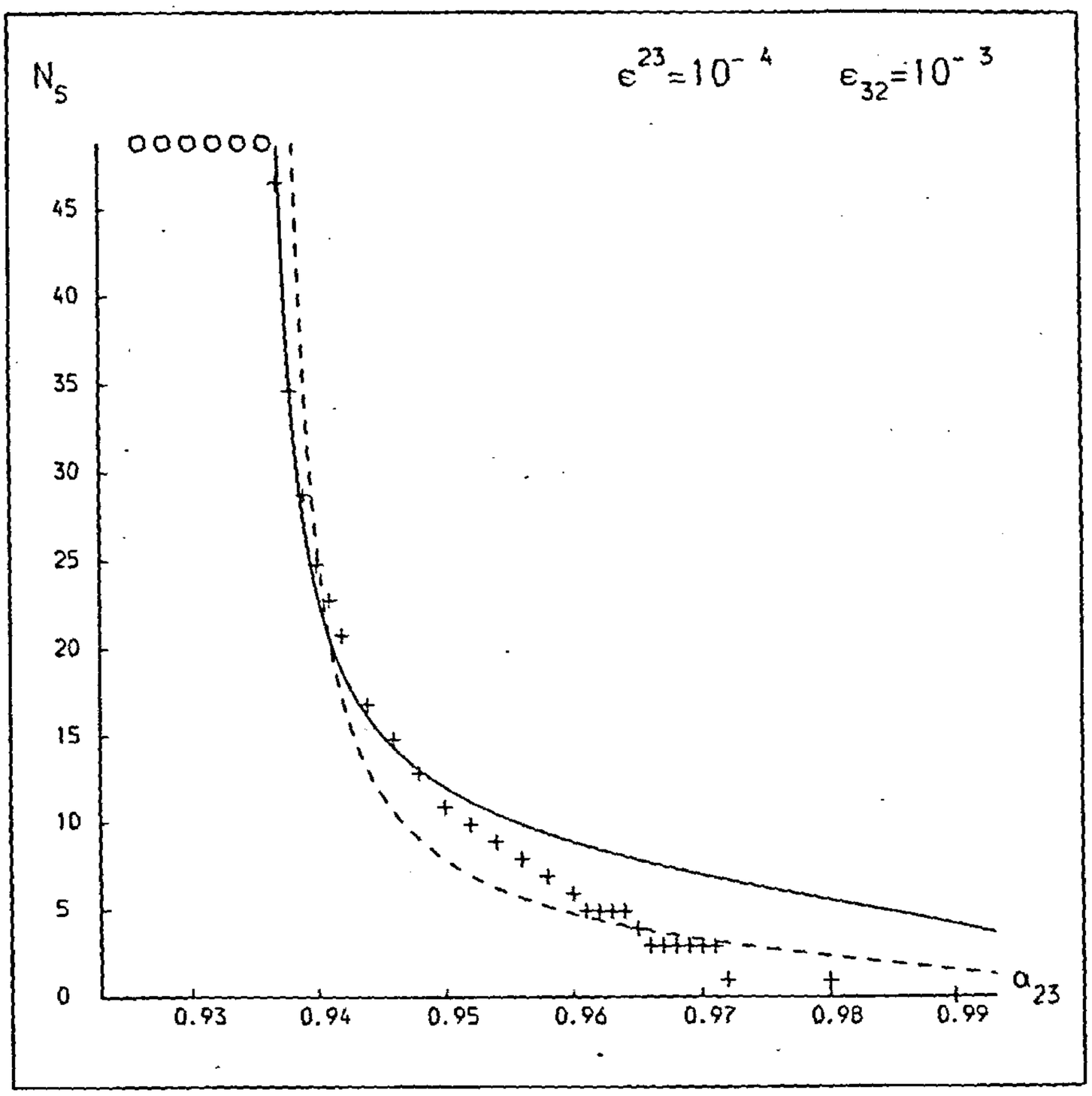

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-4}$
5
4
3
2
1
0
0.930
0.935
0.940
0.945
0.950
a_{23}

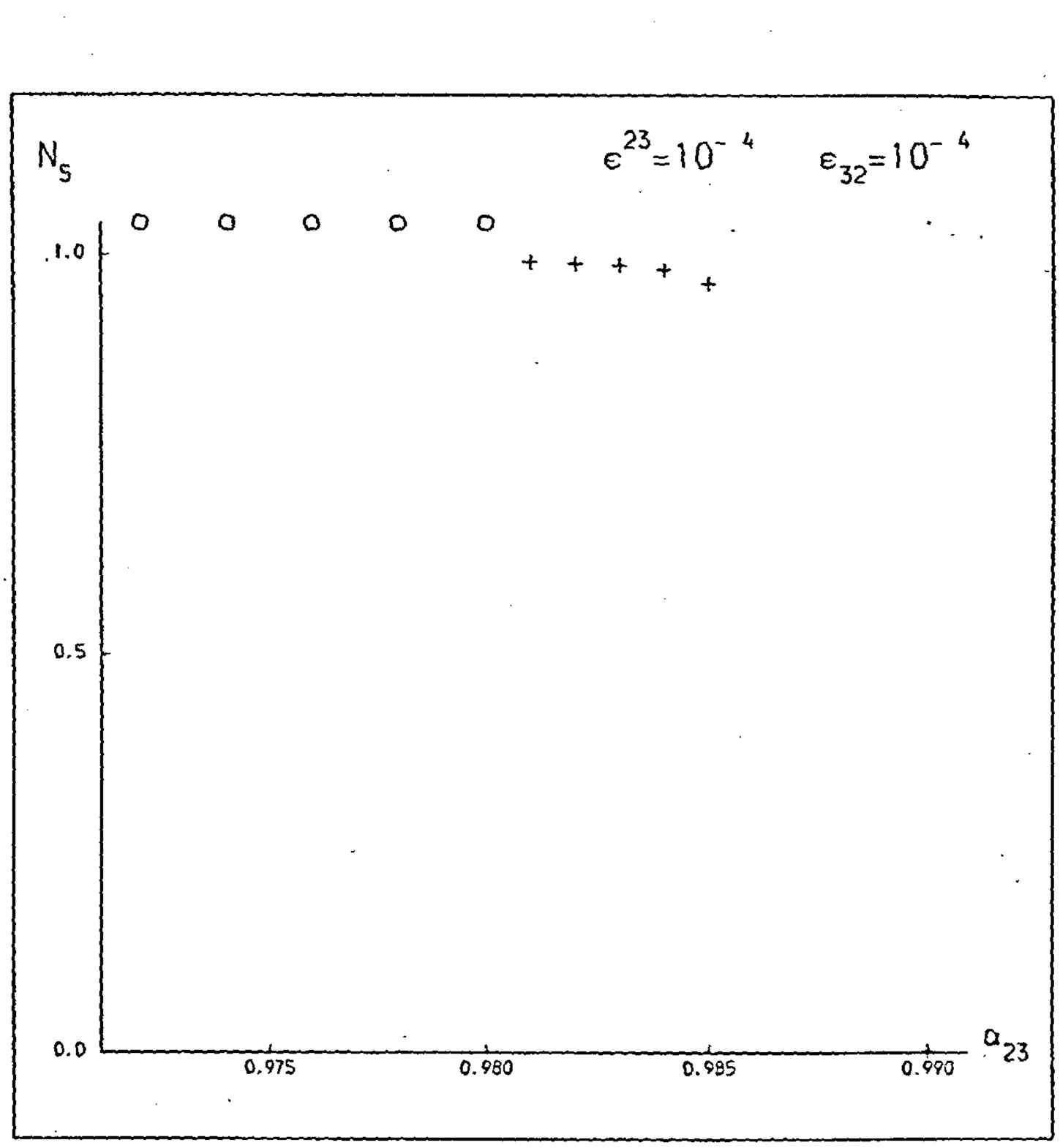

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-5}$
5
4
3
2
1
0
0.930
0.935
0.940
0.945
0.950
a_{23}

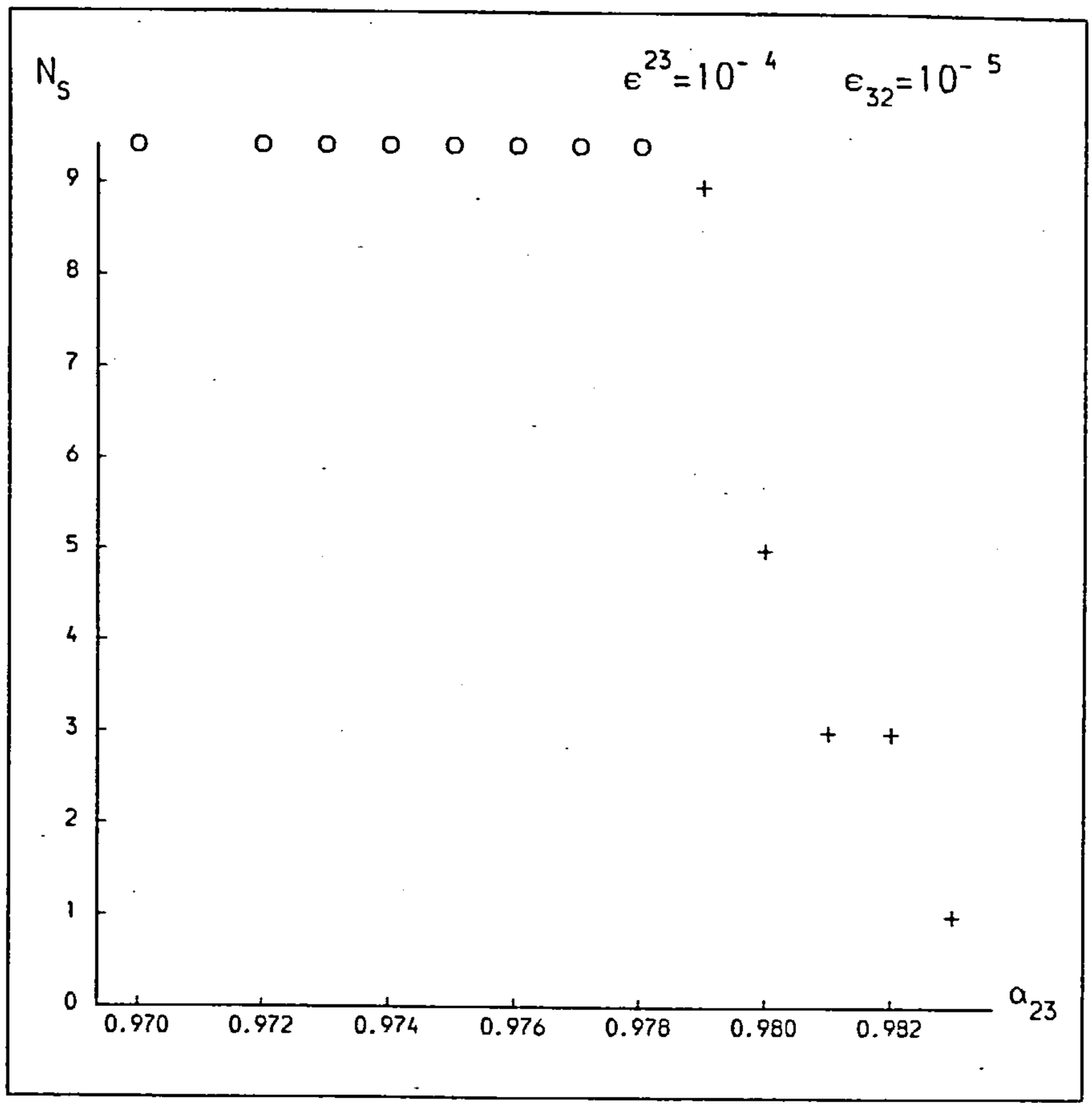

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-6}$
5
4
3
2
1
0
0.930
0.935
0.940
0.945
0.950
a_{23}

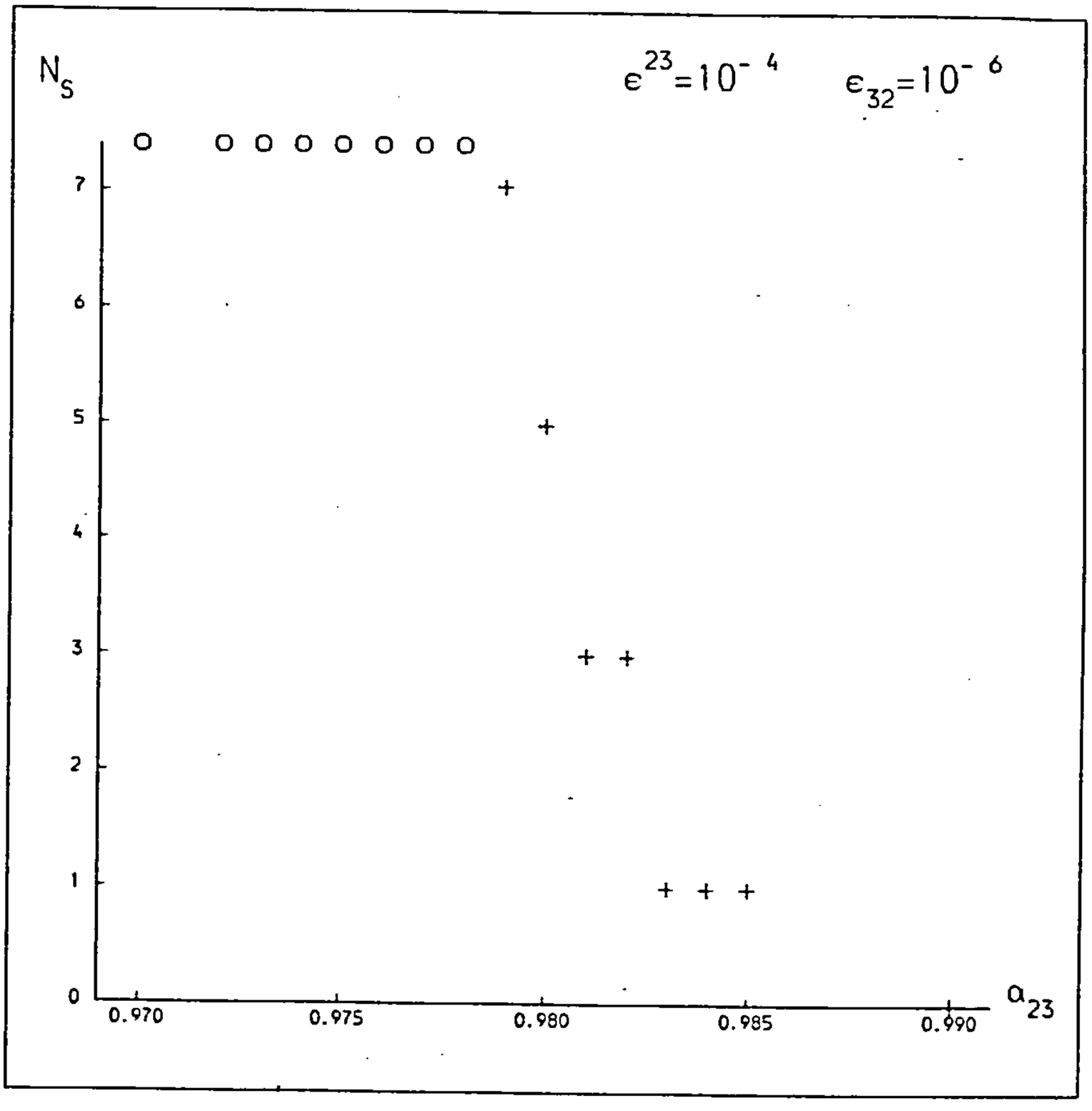

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-1}$
14
12
10
8
6
4
2
0
0.60
0.65
0.70
0.75
0.80
0.85
a_{23}

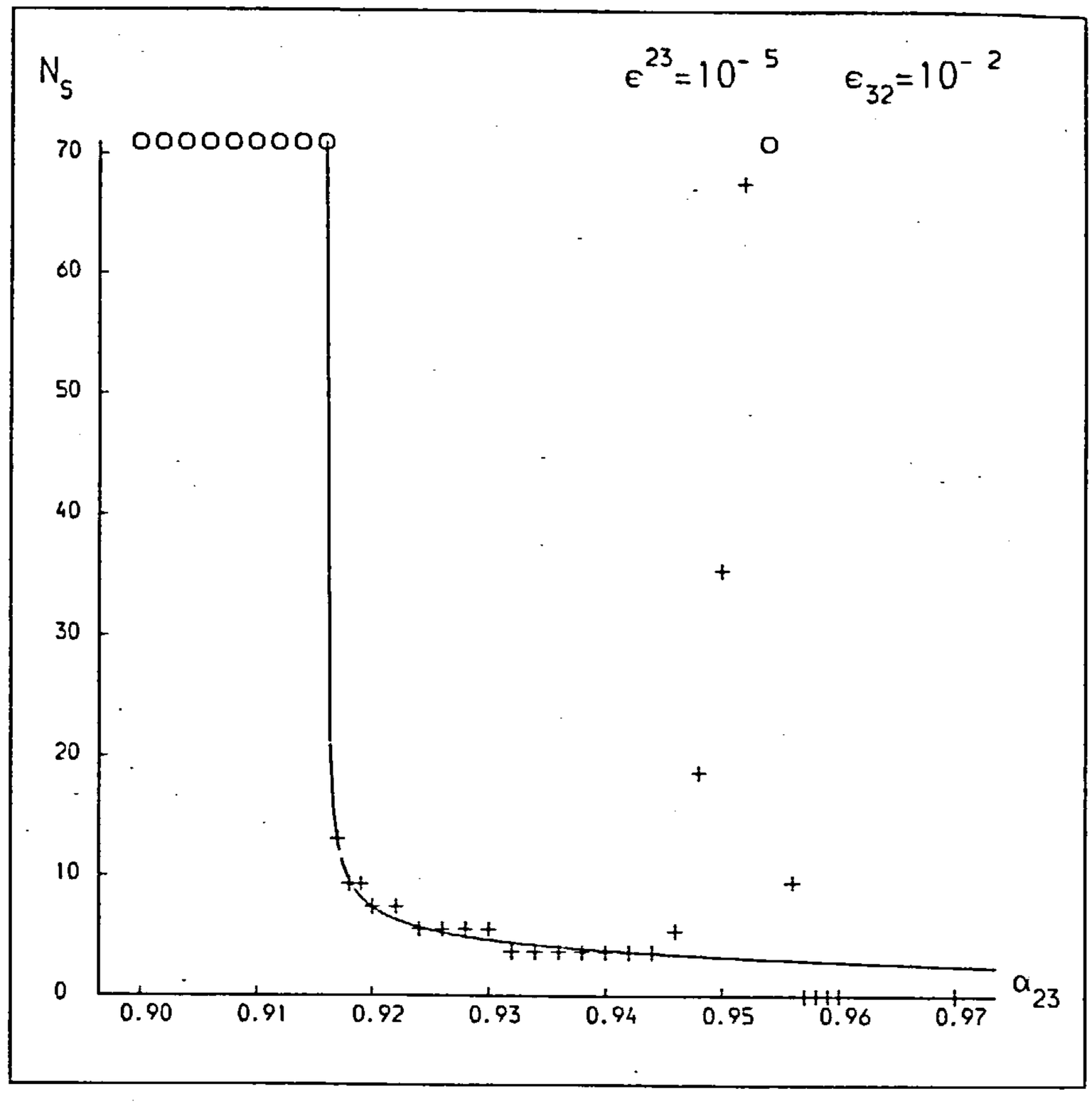

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-2}$
100
80
60
40
20
0
0.90
0.91
0.92
0.93
0.94
0.95
0.96
0.97
a_{23}

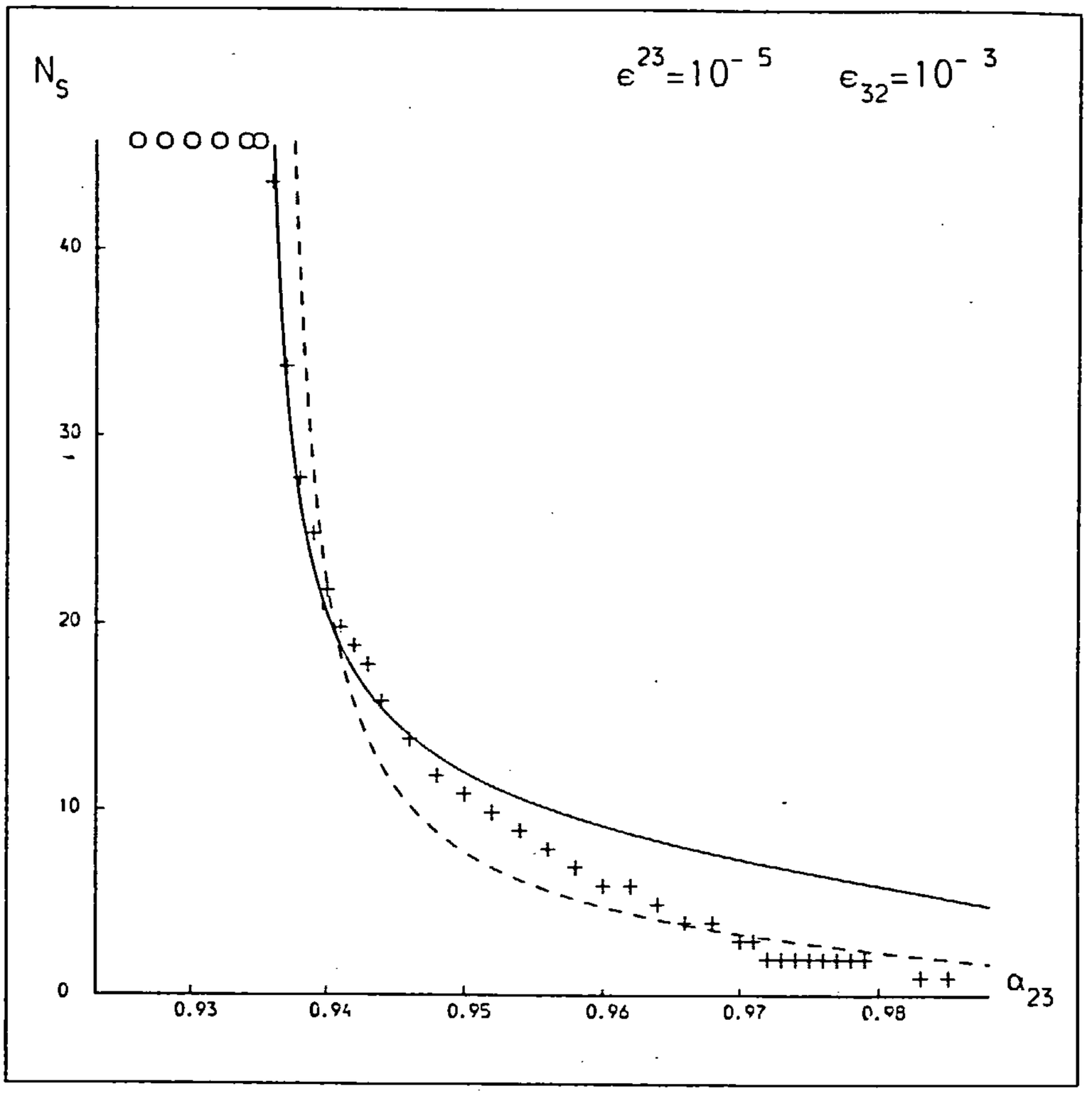

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-3}$
45
40
35
30
25
20
15
10
5
0
0.93
0.94
0.95
0.96
0.97
0.98
0.99
a_{23}

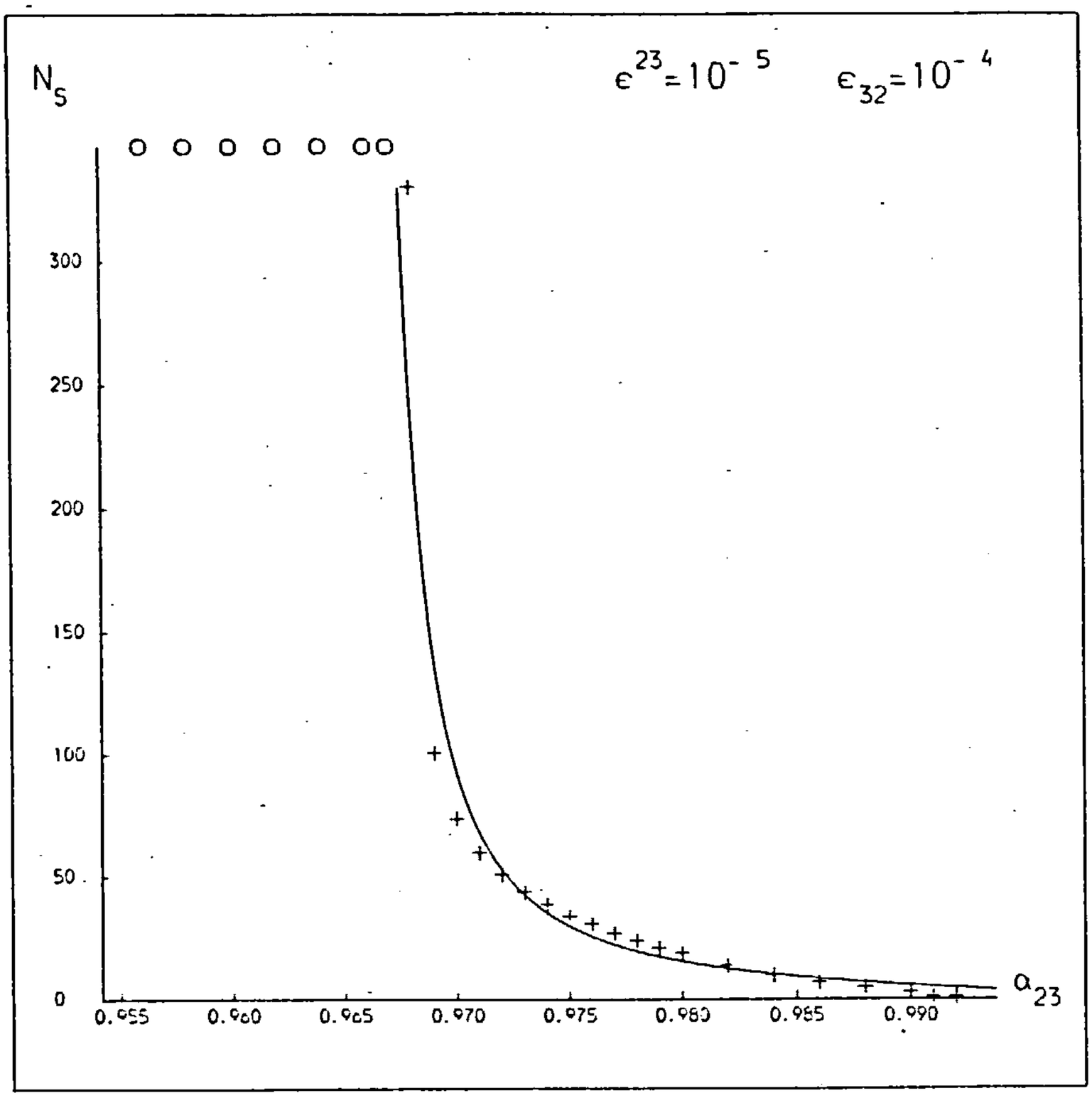

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-4}$
1.0
0.5
0.0
0.975
0.980
0.985
0.990
a_{23}

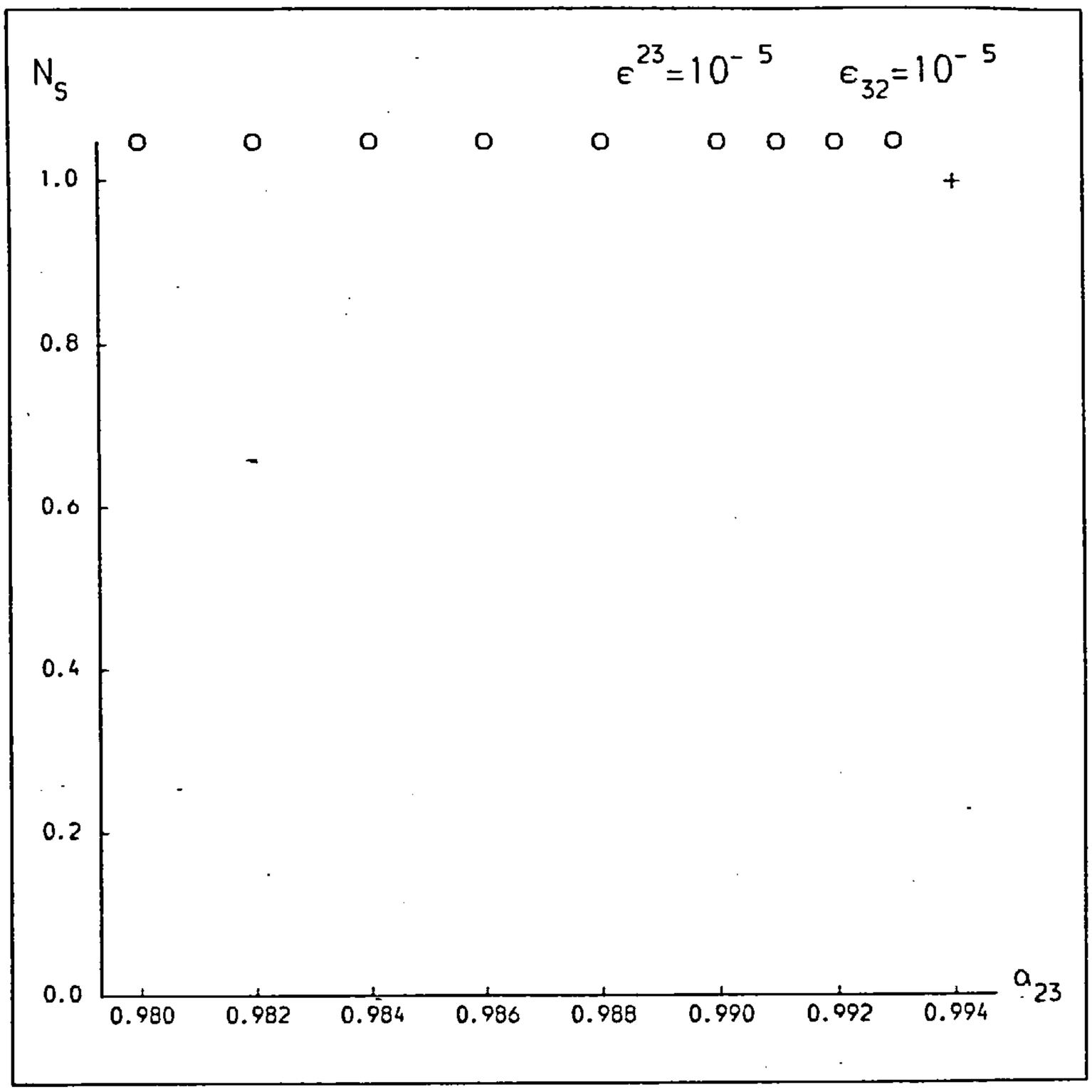

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-5}$
9
8
7
6
5
4
3
2
1
0
0.970
0.972
0.974
0.976
0.978
0.980
0.982
a_{23}

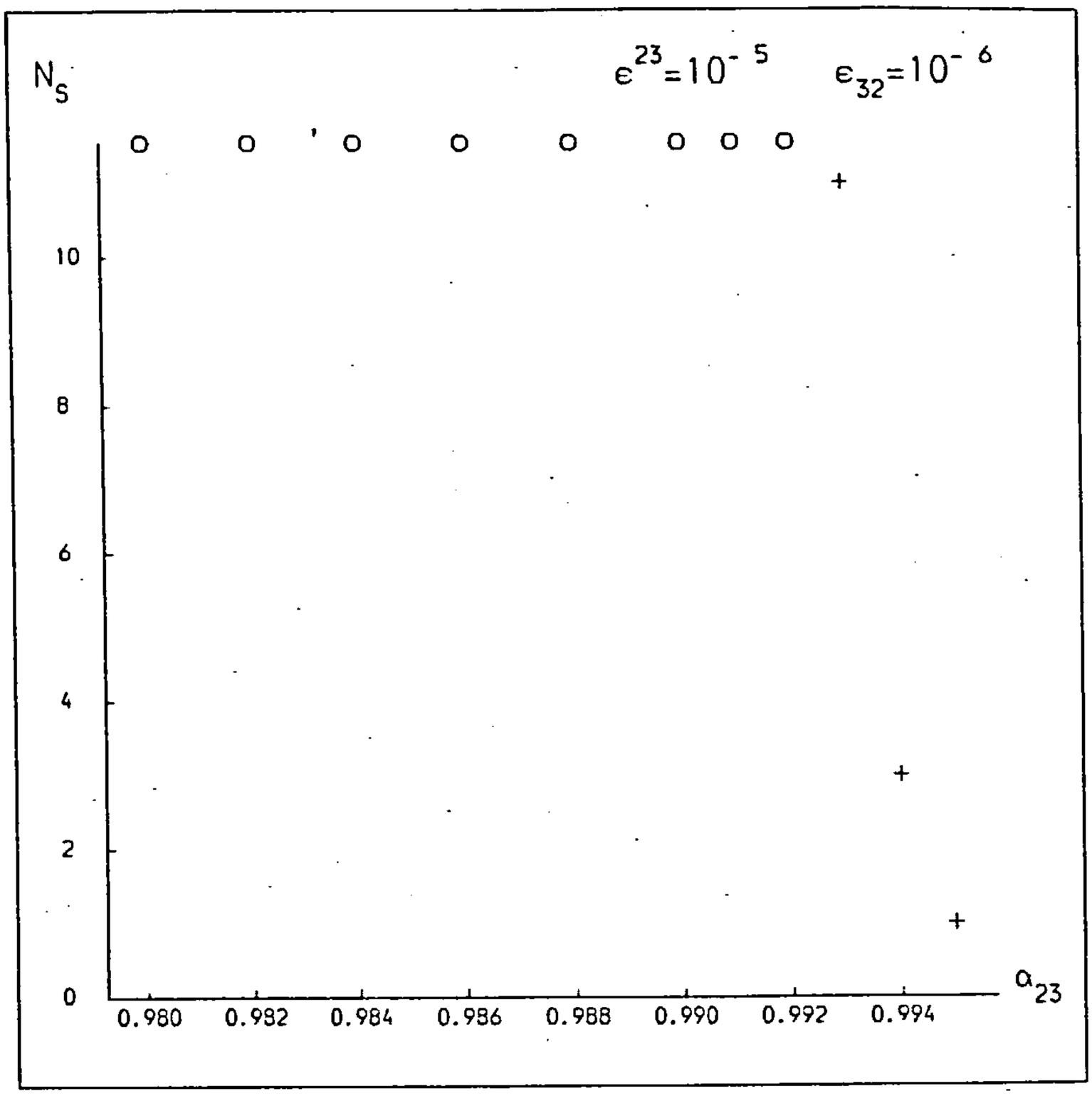

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-6}$
7
6
5
4
3
2
1
0
0.970
0.975
0.980
0.985
0.990
a_{23}

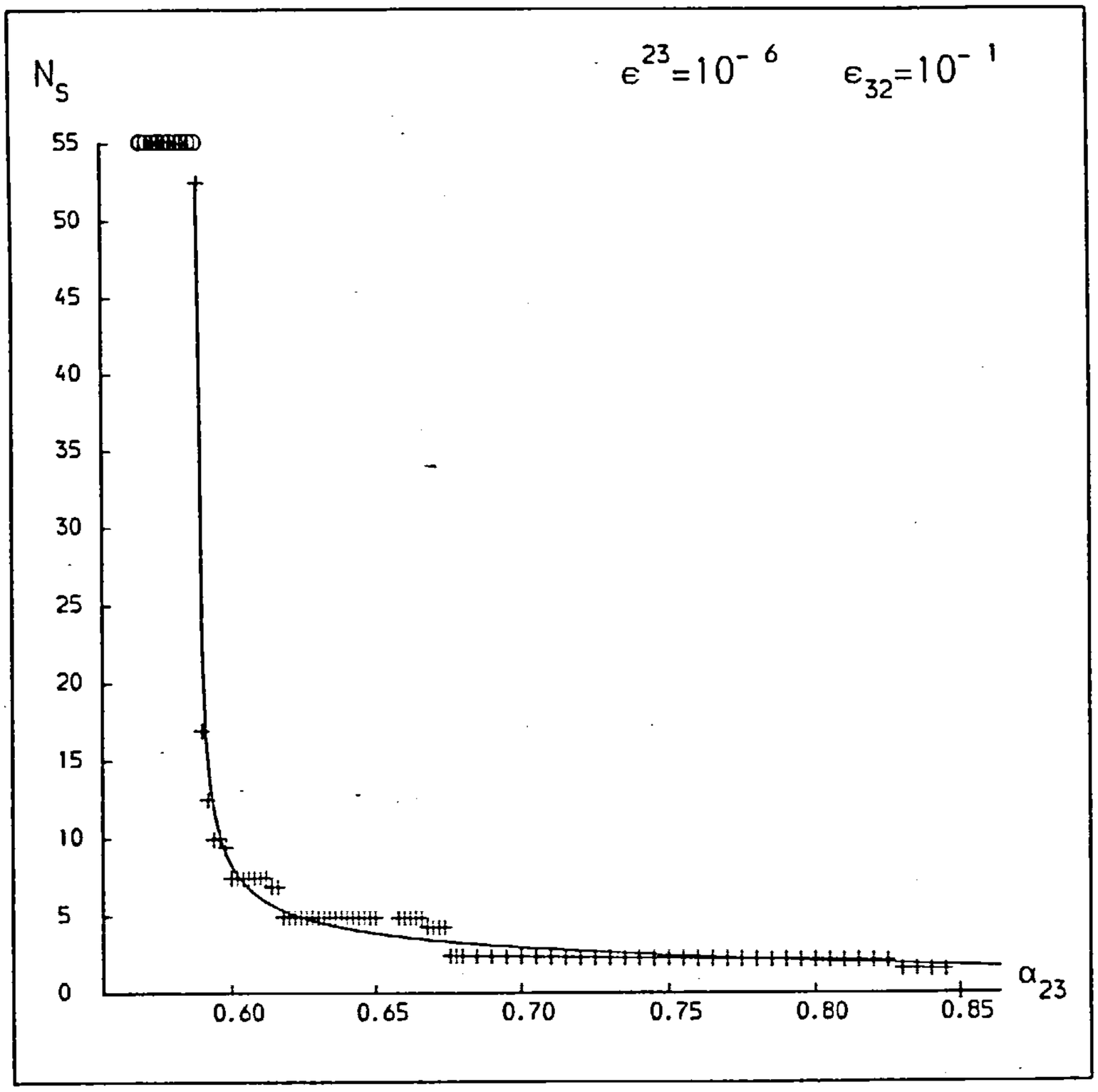

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-2}$
70
60
50
40
30
20
10
0
0.90
0.91
0.92
0.93
0.94
0.95
0.96
0.97
a_{23}

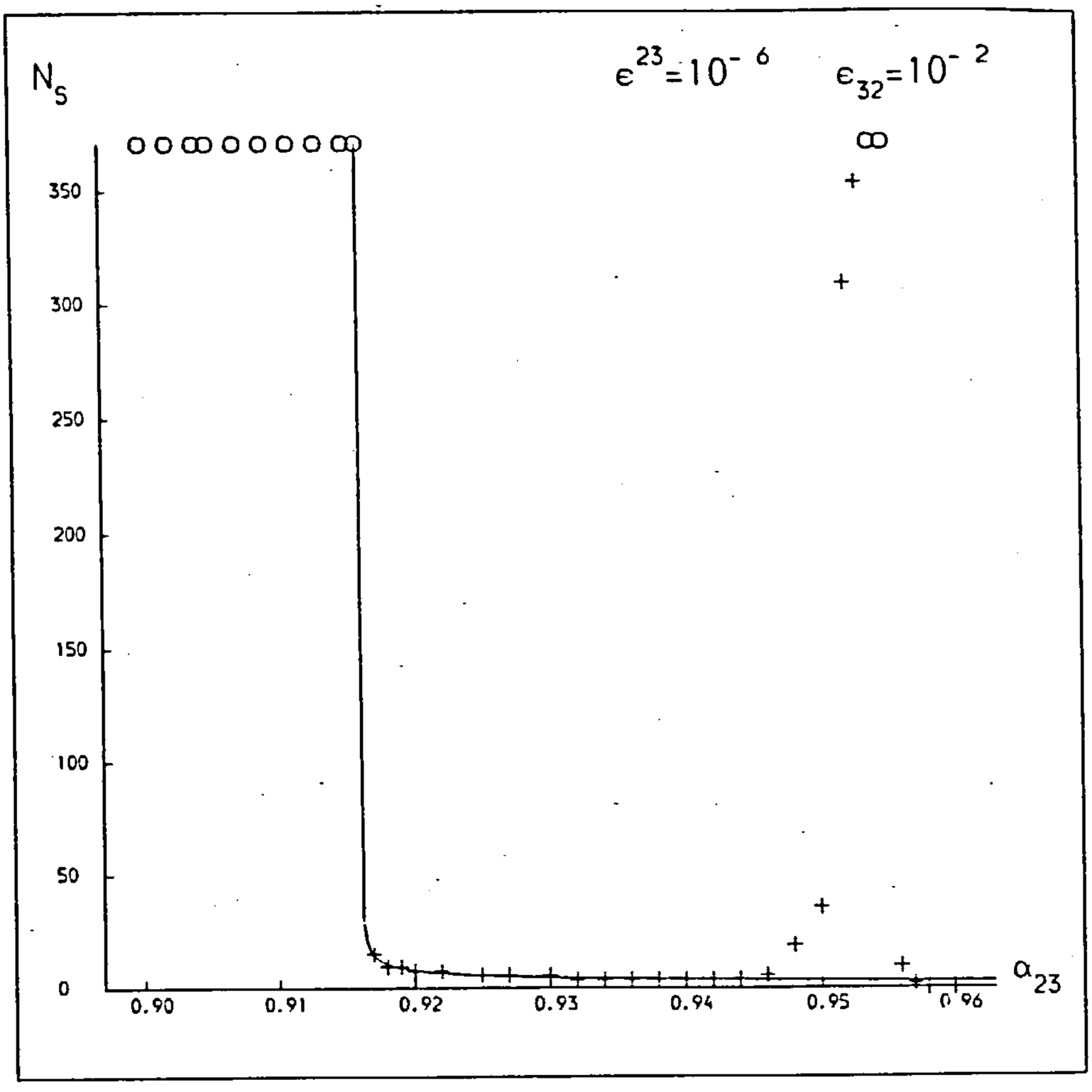

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-3}$
40
30
20
10
0
0.93
0.94
0.95
0.96
0.97
0.98
a_{23}

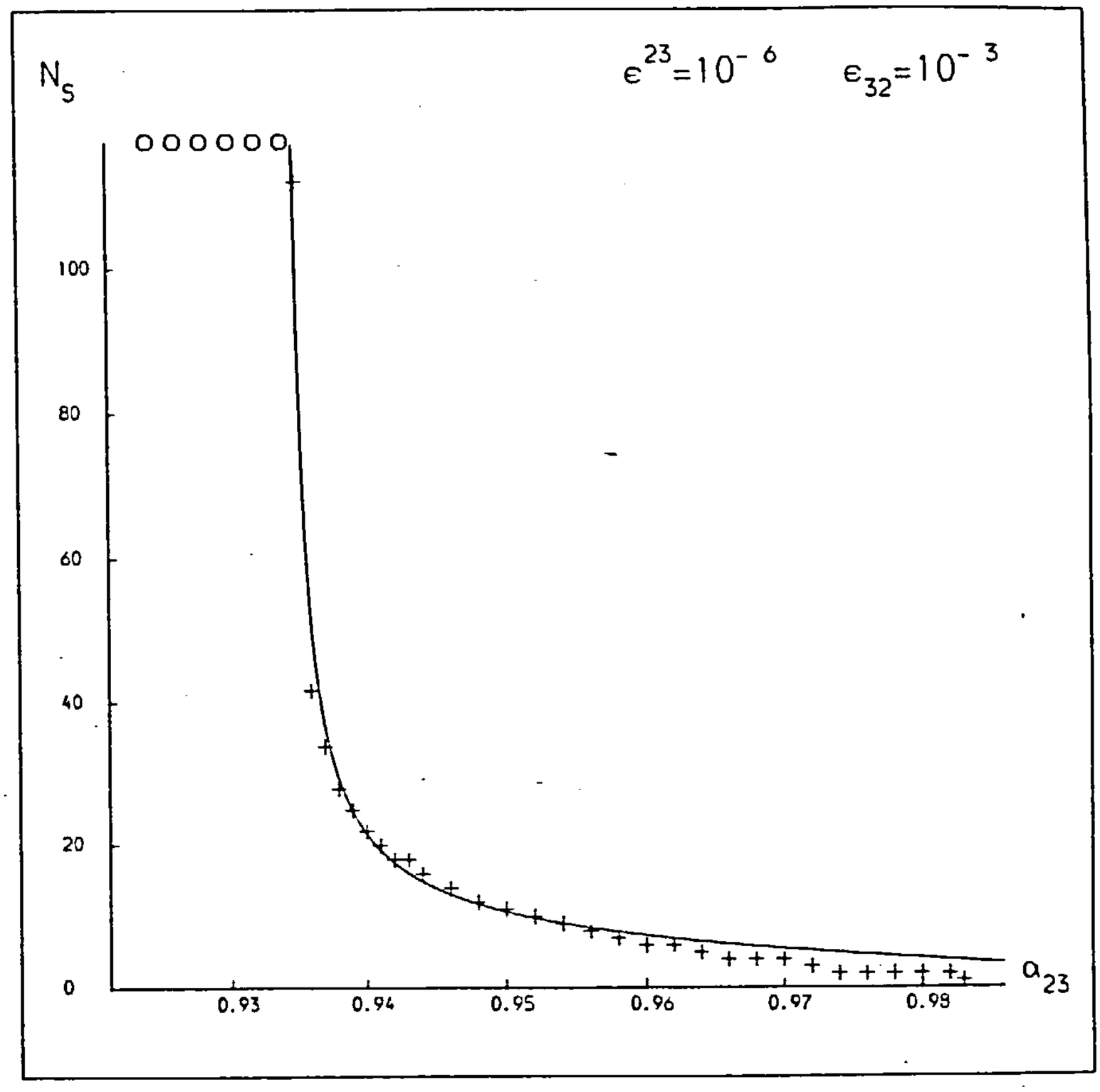

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-4}$
300
250
200
150
100
50
0
0.955
0.960
0.965
0.970
0.975
0.980
0.985
0.990
a_{23}

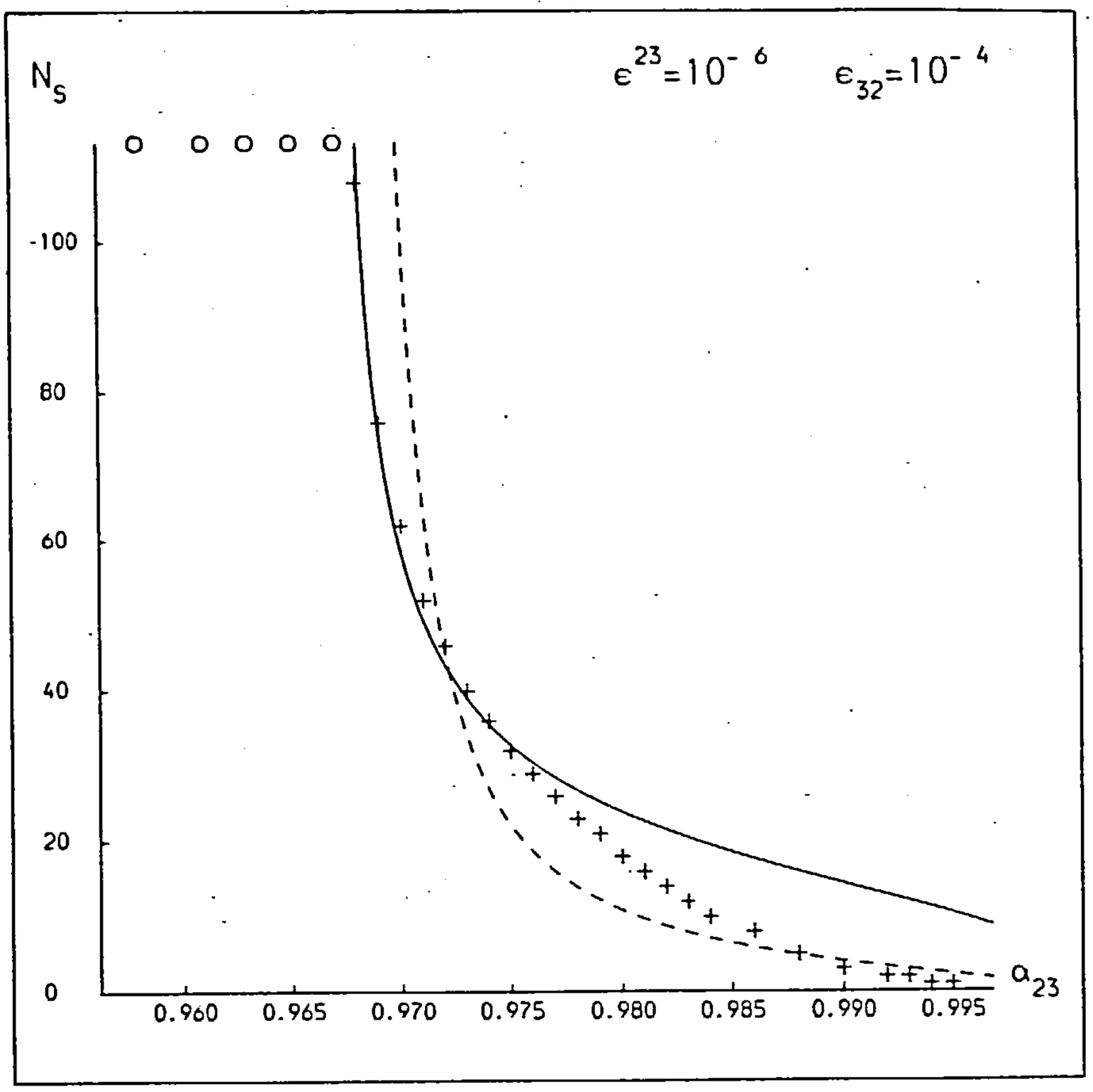

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-5}$
1.0
0.8
0.6
0.4
0.2
0.0
0.980
0.982
0.984
0.986
0.988
0.990
0.992
0.994
a_{23}

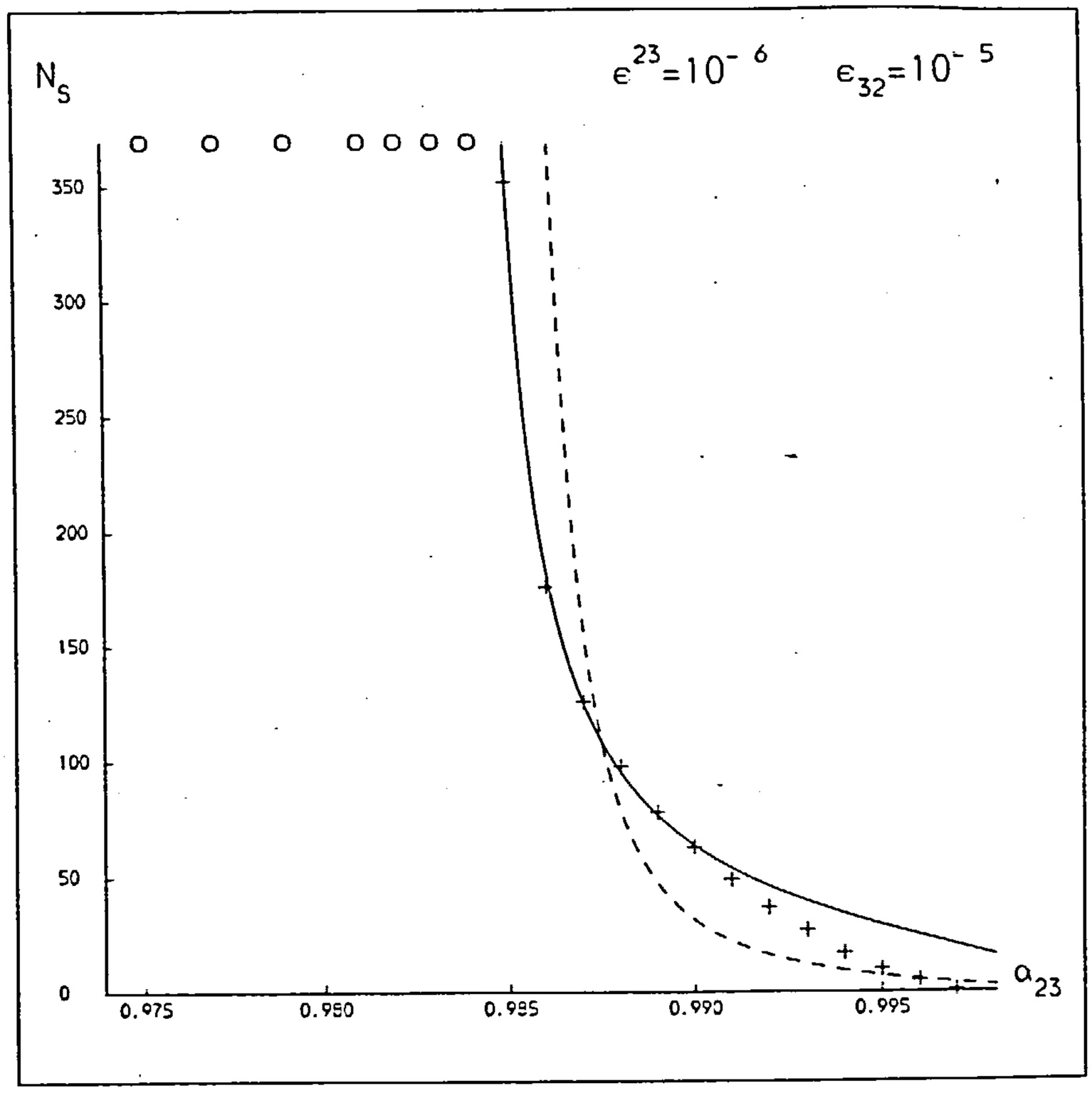

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-6}$
10
8
6
4
2
0
0.980
0.982
0.984
0.986
0.988
0.990
0.992
0.994
a_{23}

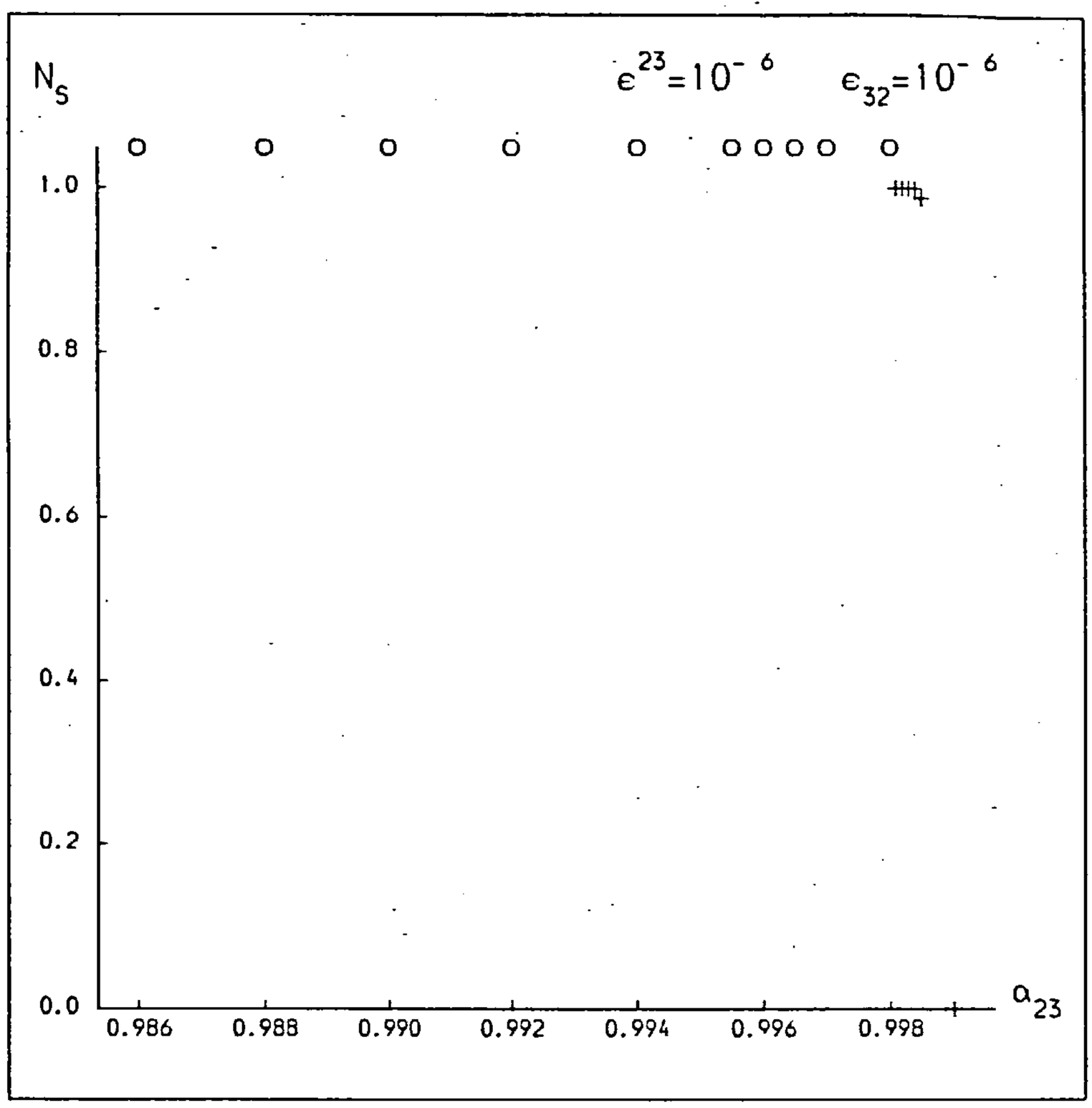

N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-1}$
55
50
45
40
35
30
25
20
15
10
5
0
0.60
0.65
0.70
0.75
0.80
0.85
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-2}$
350
300
250
200
150
100
50
0
0.90
0.91
0.92
0.93
0.94
0.95
0.96
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-3}$
100
80
60
40
20
0
0.93
0.94
0.95
0.96
0.97
0.98
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-4}$
-100
80
60
40
20
0
0.960
0.965
0.970
0.975
0.980
0.985
0.990
0.995
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-5}$
350
300
250
200
150
100
50
0
0.975
0.980
0.985
0.990
0.995
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-6}$
1.0
0.8
0.6
0.4
0.2
0.0
0.986
0.988
0.990
0.992
0.994
0.996
0.998
a_{23}

oscillate about some mean value which is proportional to ε^{23} or ε_{32}, (see Section 4.5). For a stable system, the mean value is roughly constant. For an unstable system, the mean value will rise steadily until there is an inevitable crossover of the inner and outer binary. In every case this was the observed pattern, unlike the direct systems, whose behaviours are far more varied. Because of the predictability of the retrograde systems and also because of the problem of deciding when the eccentricities had changed irreparably (the third condition for hierarchical stability given in Section 2.2), it was decided that the stability lifetime should be taken to be the time until crossover occurred.

In every graph we can see that as α decreases, the lifetimes of the systems rise, slowly at first, then extremely quickly and seem to tend to infinity at some non-zero value of α. An idea of this value can be derived from the curve-fitting procedure.

It does seem that when a system is at conjunction then it is at its least stable configuration, since the crossover always occurs around this time. There is therefore a quantisation effect with the lifetimes always being approximately integers. If a system survives one conjunction, it will survive at least until the next conjunction. We can see this in many of the graphs by noting their step-like nature.

In almost every case the stability lifetime rises monotonically as α decreases. The only exceptions are when $\varepsilon_{32} = 10^{-2}$ and $\varepsilon^{23} = 10^{-4}$, 10^{-5} or 10^{-6}. For these pairs of epsilons, there appears a blip around $\alpha = 0.95$ where the lifetimes rise for a small range in α, then drop again. Although this phenomenon is discussed again in the next section, the reason for it is not understood. The width and height

of the blip seem to increase as ϵ^{23} decreases. The value of α corresponding to the peak lifetime also seems to increase as ϵ^{23} decreases. This might indicate a possible commensurability. However, it seems unlikely as there is no strong resonance for such high values of α as 0.95. Secondly, there is no evidence of any other commensurabilities affecting the stability lifetimes in a similar way.

In each figure, it may be safely assumed that if α exceeds the highest plotted value of α for any epsilon pair, then the associated lifetime is zero, or very nearly zero due to the "integerisation" of the lifetimes.

We see three different trends depending on whether ϵ^{23} is less than, equal to, or greater than ϵ_{32}.

(a) $\epsilon^{23} < \epsilon_{32}$ The lifetimes increase gradually over a relatively large range in α. This allows us to fit a curve to the data.

(b) $\epsilon^{23} = \epsilon_{32}$ The lifetimes change from zero or one to over 600 within a vanishingly small range of α making any curve fit impossible and totally unnecessary. All that need be quoted is the asymptote α_o.

(c) $\epsilon^{23} > \epsilon_{32}$ The lifetimes increase very quickly but not as quickly as when $\epsilon^{23} = \epsilon_{32}$. However there is still little point in fitting any curve to such sparse data as is shown in Figures 4.1.

When $\epsilon^{23} \geq \epsilon_{32}$, one need only quote the observed asymptotic cut off α_o, and be reasonably sure that any system with $\alpha < \alpha_o$ will be stable for over 600 synodic periods (and quite possibly for all time). If $\alpha > \alpha_o$, the system is highly unstable and will usually only survive for a few synodic periods (< 10).

The curve-fitting procedure now described is only applied to systems with $\epsilon^{23} < \epsilon_{32}$. We recall from Section 4.2 that the lifetime

is infinite for $\alpha < \alpha_o$, by definition of α_o, and is zero for $\alpha > \alpha_x = 1/(1-\mu) \simeq 1$. We must therefore find a suitable curve for modelling the stability lifetimes N_s as a function of $\varepsilon^{23}, \varepsilon_{32}$, α, $\forall \alpha \in (\alpha_o, 1]$. It should be continuous and monotonically decreasing with α. We are concerned with fitting curves to individual graphs among Figures 4.1. Hence we fit curves of the form

$$N_s = f(\alpha;\ c_1,\ c_2, \ldots, c_k)$$

where $k \in \mathbb{N}$ and $c_i = c_i\ (\varepsilon^{23}, \varepsilon_{32})$, $\forall i = 1, \ldots k$. Clearly if $k > 2$, the parameters c_i cannot be independent. The more parameters we include in function f, the better the fit is likely to be. Obviously it would be desirable for simplicity to limit the number of parameters. For example a step function would fit the data well but would require many parameters to denote the length and height of each step.

Walker (1980) investigated two choices of function for direct systems:

$$\text{(a)}\ N_s = f_1(\alpha;\alpha_o,\gamma) = \left(\frac{1-\alpha}{\alpha-\alpha_o}\right)^{\gamma}, \quad \alpha_o < \alpha \leqslant 1. \qquad (1)$$

$$\text{(b)}\ N_s = f_2(\alpha;\alpha_o,\beta,\gamma) = \exp\left[\ \beta\left(\frac{1-\alpha}{\alpha-\alpha_o}\right)^{\gamma}\ \right] - 1, \quad \alpha_o < \alpha \leqslant 1 \qquad (2)$$

Both curves are monotonically decreasing, $N_s \to \infty$ as $\alpha \to \alpha_o$, and $N_s = 0$ at $\alpha = 1$. Walker considered that f_1, with two free parameters, α_o and γ, was not sufficiently pliable in order to fit the data accurately enough. He preferred f_2 with three free parameters α_o, β and γ. To have any more was considered unnecessary because of the noisiness of the data.

Having chosen the functional form of the curve, it is now necessary to find the best fitting values of α_o, β, γ for any given ε^{23}, ε_{32} from the data at our disposal. We may re-express Equation (2) as

4.4

$$\log(\log(N_s + 1)) = \log \beta + \gamma \log \left(\frac{1-\alpha}{\alpha-\alpha_o}\right) \qquad (3)$$

Consider n systems with common $\varepsilon^{23}, \varepsilon_{32}$, where α_i is the initial ratio of semi-major axes and $(N_s)_i$ the stability lifetime of the i^{th} system. One can evaluate the quantities u_i, v_i, where

$$u_i = \log \left(\frac{1-\alpha_i}{\alpha_i-\alpha_o}\right) \qquad i = 1,\ldots,n \qquad (4)$$

$$v_i = \log(\log((N_s)_i + 1)) \qquad (5)$$

having chosen a test value of α_o. If Equation (2) accurately describes the distribution, then the (u_i, v_i) data should lie on a straight line with gradient γ and intercept at $\log \beta$. A least squares fit is chosen to give the best fit value of β and γ for a given α_o. Walker chose by eye the value of α_o (with its associated β,γ) which looked like it gave the best straight line. A more objective way is to calculate the value of the correlation coefficient for each set of α_o, β, γ chosen.

The correlation coefficient r is a useful statistical tool for examining data correlation and goodness of fit. Details can be found in many books on the statistical analysis of data (eg. Anderson and Sclove). In this case

$$r^2 = \frac{\left[\sum_{i=1}^{n} (u_i-\bar{u})(v_i-\bar{v})\right]^2}{\sum_{i=1}^{n} (u_i-\bar{u})^2 . \sum_{i=1}^{n} (v_i-\bar{v})^2} \qquad (6)$$

where

$$n\bar{u} = \sum_{i=1}^{n} u_i \quad , \quad n\bar{v} = \sum_{i=1}^{n} v_i$$

4.4

If r=1, then the (u_i,v_i) data is arranged in a perfectly straight line. The worse the fit, the lower r becomes. If there is no correlation at all, then r=0. Obviously we choose the value of α_o which maximises r and compute β and γ from it.

It is considered to be more important to get a good fit for the low α-high N_s data than for the high α-low N_s data. There is little point in knowing accurately the lifetime of a system which we know will not survive for more than a few synodic periods. For systems where α is very close to α_o, an inaccurate curve may produce an error of tens or even hundreds of synodic periods. For this reason, the least squares fit is normally weighted in favour of the low α-high N_s data by introducing a weighting factor proportional to the lifetime of each system.

There are many more data points per "step" for low lifetime systems which will tend to weight a fit in favour of the very unstable systems. To counteract this, only the middle data point in each lifetime step is considered in fitting the curve. For example, for $\varepsilon^{23} = 10^{-4}$ and $\varepsilon_{32} = 10^{-3}$, there are six systems with lifetimes approximately equal to three synodic periods. Only one point near the middle is used for curve fitting, namely $\alpha = 0.969$. This will reduce the available data set, but increase the accuracy of the fit.

Although some systems show lifetimes exceeding 600 synodic periods, this does not necessarily mean that they are stable for all time. It may be that α_o is substantially less than α_{us}, the lowest value of α for which an unstable system is detected. Any best fit curve should however reflect the fact that the systems circled in Figures 4.1 will have lifetimes greater than 600 synodic periods, even if their α value

4.4

exceeds a chosen α_o. For example, if the value of α_o is too low, the best fit curve may predict that a circled system will survive for 100 synodic periods. This particular value of α_o should therefore be rejected.

By restricting α_o in this way as well as using the correlation coefficient, it is possible to describe $N_s = N_s(\alpha)$ for any $\epsilon^{23}, \epsilon_{32}$ in a completely objective manner. This was done for all $\epsilon^{23} < \epsilon_{32}$. The results are given in Table 4.1 for weighted and unweighted least squares fits. For those epsilon pairs that exhibit an anomalous peak in the lifetimes, the curves are fitted without using the data that make up the behaviour. Table 4.1 gives the critical values of α, i.e. α_c, α_o, α_{us}. The analytical value of α for stability, α_c, does not always exist and when it does is substantially lower than the empirical stability limit, α_o. Note that α_o is always very close to α_{us}.

For systems with $\epsilon^{23} < \epsilon_{32}$, curve-fitting parameters β, γ are tabulated alongside the correlation coefficient r.

The weighted least squares fit is preferred, as stated earlier. However in some cases the discrepancy between this curve and the data is quite marked for high α-low N_s data. Thus the unweighted fit is included for comparison and drawn with a dotted line in Figures 4.1. Anyone wishing to use these curves for determining the lifetime of an unknown system may choose the more appropriate curve.

The reason why we sometimes need both is best explained by looking at the (u,v) data and corresponding best fit straight lines for one pair of epsilons. Figure 4.2 is such a graph for $\epsilon^{23} = 10^{-5}$, $\epsilon_{32} = 10^{-4}$. The full and dotted lines are the weighted and unweighted best fit straight lines respectively. Note that the points are weighted for

4.4

ε_{23}	ε_{32}	α_c	α_o	α_{us}	Weighted β	Weighted γ	Weighted r	Unweighted β	Unweighted γ	Unweighted r
10^{-2}	10^{-1}	-	0.553	0.554	1.548	0.111	0.83			
	10^{-2}	-	0.822	0.823						
	10^{-3}	-	0.834	0.835						
	10^{-4}	-	0.834	0.835						
	10^{-5}	-	0.834	0.835						
	10^{-6}	-	0.834	0.835						
10^{-3}	10^{-2}	-	0.901	0.903	0.502	0.566	0.95			
	10^{-3}	0.145	0.938	0.939						
	10^{-4}	-	0.938	0.939						
	10^{-5}	-	0.938	0.939						
	10^{-6}	-	0.938	0.939						
10^{-4}	10^{-1}	-	0.589	0.590	1.574	0.098	0.82			
	10^{-2}	-	0.913	0.914	0.713	0.418	0.94			
	10^{-3}	0.203	0.936	0.937	2.118	0.151	0.68	1.510	0.288	0.86
	10^{-4}	0.180	0.980	0.981						
	10^{-5}	0.081	0.978	0.979						
	10^{-6}	-	0.978	0.979						
10^{-5}	10^{-2}	-	0.916	0.917	1.364	0.148	0.99			
	10^{-3}	0.212	0.935	0.936	2.170	0.140	0.70	1.528	0.290	0.92
	10^{-4}	0.233	0.965	0.968	2.558	0.320	0.86			
	10^{-5}	0.188	0.993	0.994						
	10^{-6}	0.084	0.992	0.993						
10^{-6}	10^{-1}	-	0.588	0.589	1.131	0.193	0.98			
	10^{-2}	-	0.916	0.917	1.330	0.162	0.94			
	10^{-3}	0.212	0.934	0.935	1.932	0.207	0.87			
	10^{-4}	0.241	0.967	0.968	3.041	0.128	0.55	2.136	0.338	0.89
	10^{-5}	0.239	0.984	0.985	3.852	0.155	0.54	2.848	0.391	0.84
	10^{-6}	0.189	0.997	0.998						

Table 4.1: Summary of critical values of α obtained from the results displayed in Figures 4.1. Curve fit parameters and correlation coefficients are also given for the cases where weighted and unweighted curves are drawn through the data. These results apply for retrograde systems only.

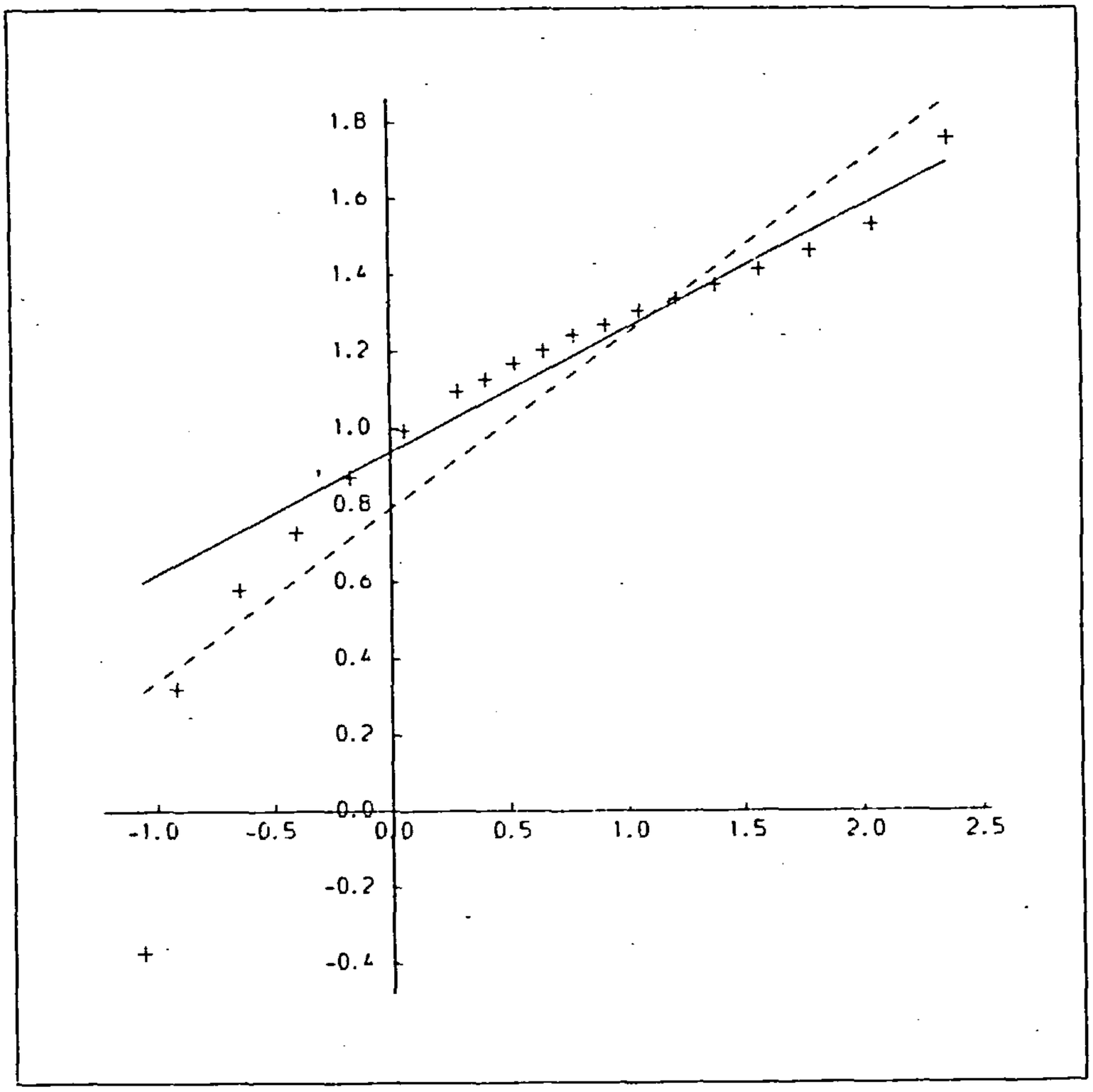


Figure 4.2: v against u for $\epsilon^{23} = 10^{-5}$. $\epsilon_{32} = 10^{-4}$. The crosses represent the real data. The full and dotted lines are the weighted and unweighted least squares fits respectedly.

high u and v which corresponds to low α-high N_s (Equations (4) and (5)).

It is now obvious that there is a systematic error in the curve. Moreover, it is not an error that we can eliminate easily. The (u,v) points which should lie in a straight line, in fact form an "S"-shape. When fitting the curves for various α_o, if we choose a high value of α_o the bending at high u,v can be eliminated, but only at the cost of making the bending at low u,v worse. Alternatively, for low α_o, the bending at low u,v is curable at the cost of increasing the bending at high u,v. We prefer to accept the former for the reasons expressed previously and a weighted fit will hence tend to lead to a higher value of α_o. The unweighted fit will tend to seek a compromise between the two extremes, hence its correlation coefficient is normally higher than its weighted counterpart. There is little that can be done to remedy this error, except choose a different functional form for $N_s = f(\alpha)$. Other functions will be mentioned in Chapter 8. However, this function is generally adequate for determining lifetimes and if used properly should only lead to small errors in the lifetimes that it predicts.

Returning to Table 4.1, we notice some trends. α_o is almost monotonic as it increases, while ε^{23} and ε_{32} decrease. There is some evidence that the same is true for β. There is however less of a trend for γ. It may be that we are looking for trends that are not present, since β and γ may be very complicated functions of ε^{23} and ε_{32}. A more likely reason could be that the errors in the fitted value of γ (and possibly β) are so large as to make any trend difficult to see.

The difference between α_c and α_o will be discussed at greater length in Chapter 6.

4.5 On the General Behaviour of Retrograde Three-Body Systems

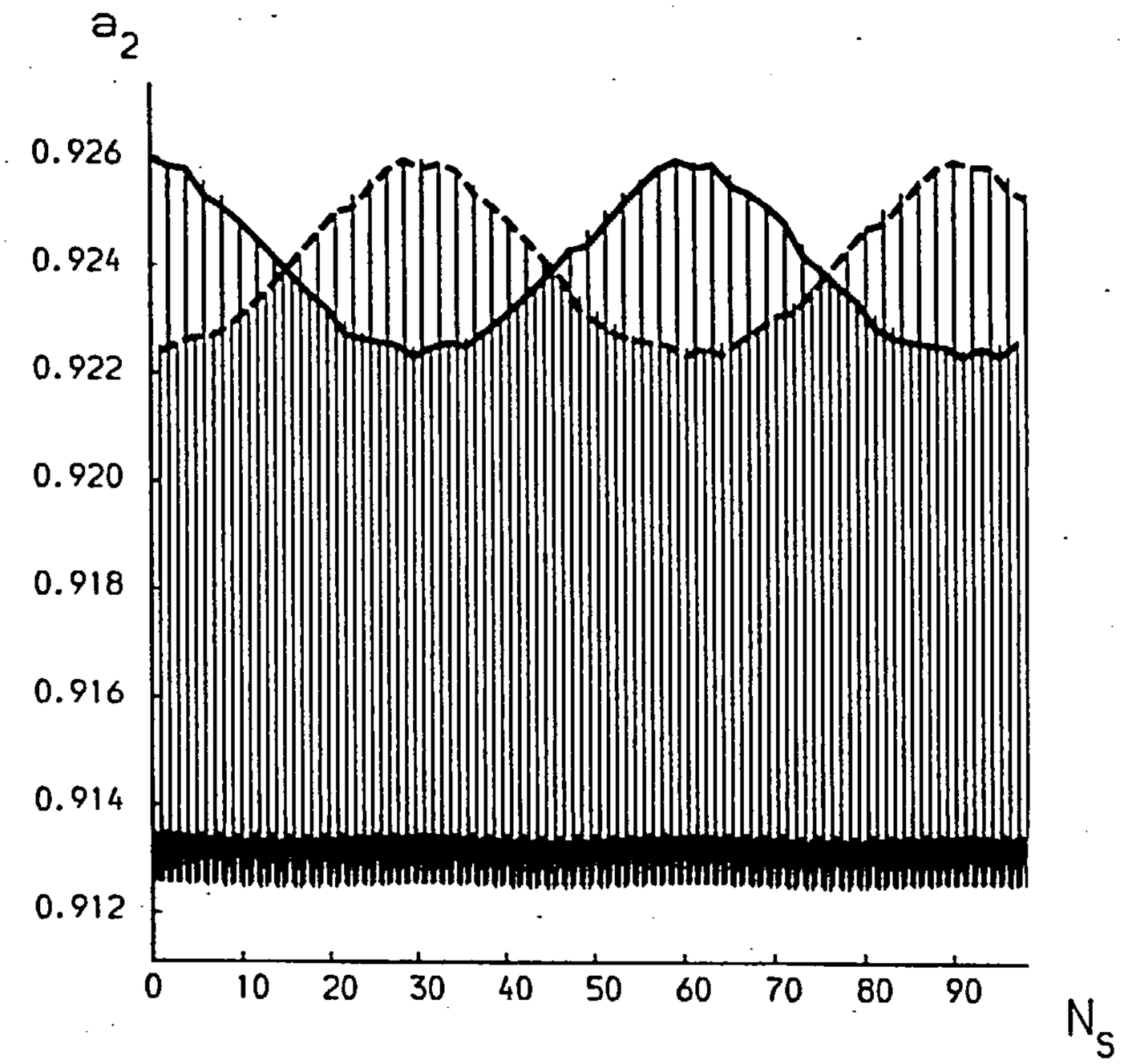
The trends discussed in the previous section mostly pertained to groups of systems with common ε^{23} and ε_{32} . In this section we shall investigate characteristic behaviour that holds for all systems. These characteristics will be highlighted by particular examples taken from the dataset of systems that were numerically integrated in Section 4.4. These systems were re-integrated and the elements a and e were plotted step by step for each binary orbit in the system. These plots are given in Figures 4.3. Note that at time zero, $a_2 = \alpha$, $a_3 = 1$, $e_2 = e_3 = 0$. There is usually little qualitative difference between a_2 and a_3, or e_2 and e_3. Therefore on a_2 and e_2, the change in elements every conjunction are superimposed on the changes every step. For reasons that will become apparent two such curves are drawn. The full one graphs a_2 or e_2 against time (measured in synodic periods N_s) at every even conjunction, (i.e. at the zeroth, 2nd, 4th, 6th,... conjunction). We denote it by subscript e, (eg. $(a_2)_e$). The dotted curve graphs a_2 or e_2 against time at every odd conjunction (i.e. at the 1st, 3rd, 5th, ... conjunction), denoted by subscript o, (eg. $(a_2)_o$). There is a lot of information contained in each graph presented here, and in others like them. All that is done in this section is to point out particularly interesting features which may or may not have a direct bearing on the stability of the system in question. Suggestions for a more detailed analysis are given in Chapter 8.

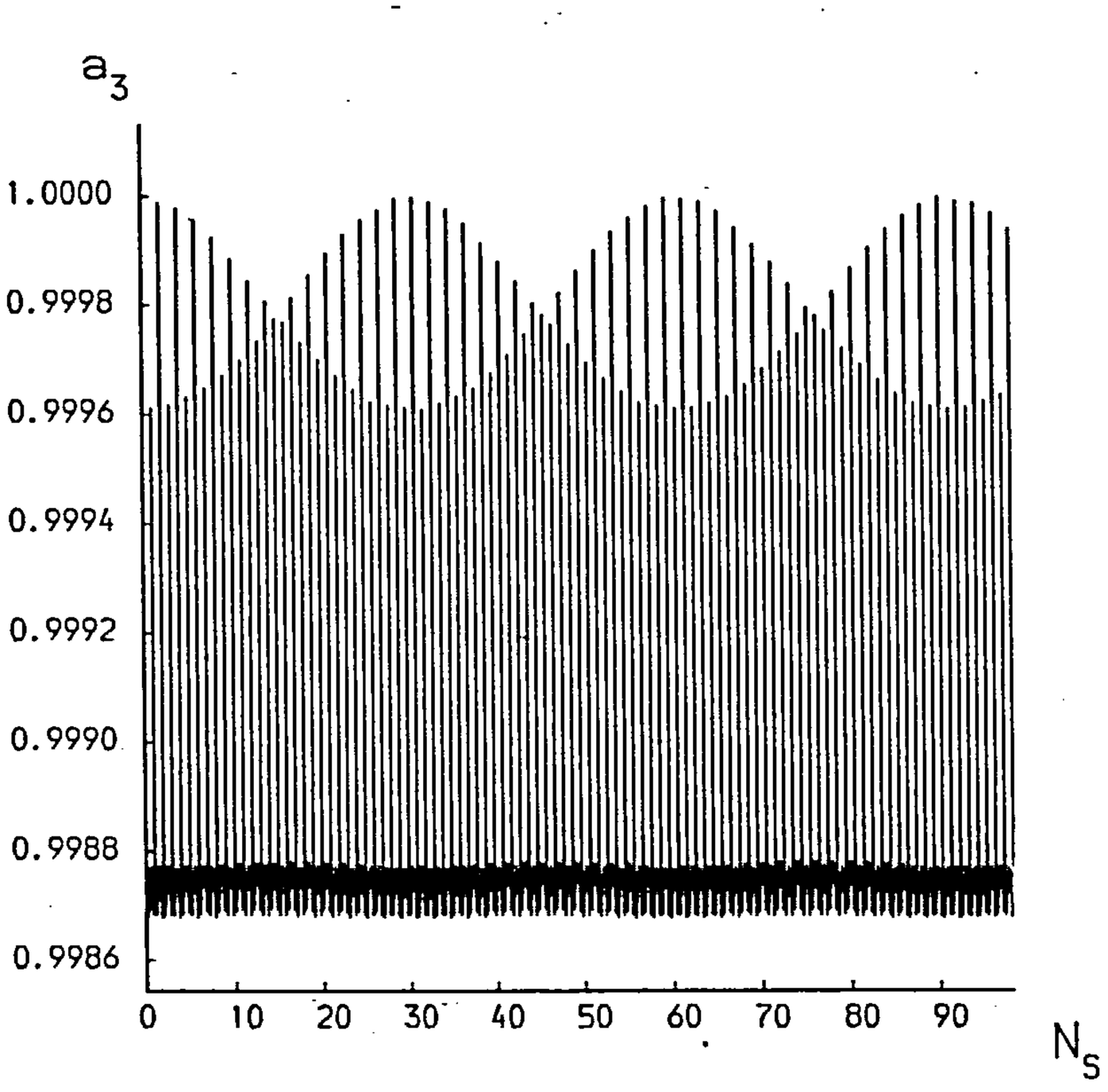
In general, the maximum values of a_2 and a_3 per synodic period occur at conjunction showing a strong peak before falling back sharply to a lower more constant value (Figure 4.3(b)). There is a maximum at opposition but it is usually very much smaller than the maximum

Figures 4.3

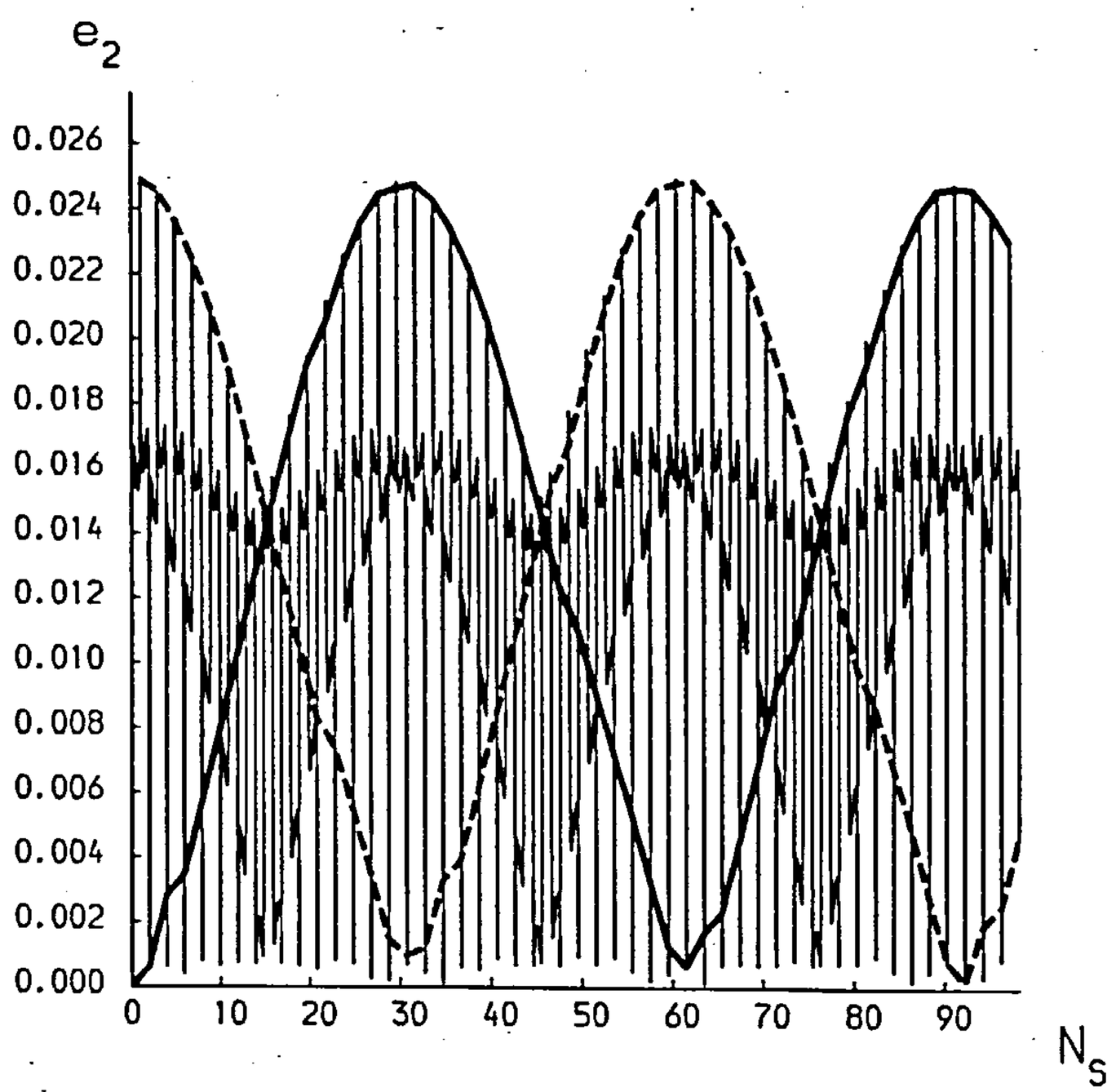
On the following pages there are graphs of a_2, a_3, e_2, e_3 against time (N_s) for various systems listed below. The thin line indicates the variation every step. The thick line indicates the variation at every even conjunction; the dotted line indicates the variation at every odd conjunction (for a_2, e_2 only). The unstable systems are graphed over their whole stability lifetimes. The stable systems are only graphed over the first 100 conjunctions.

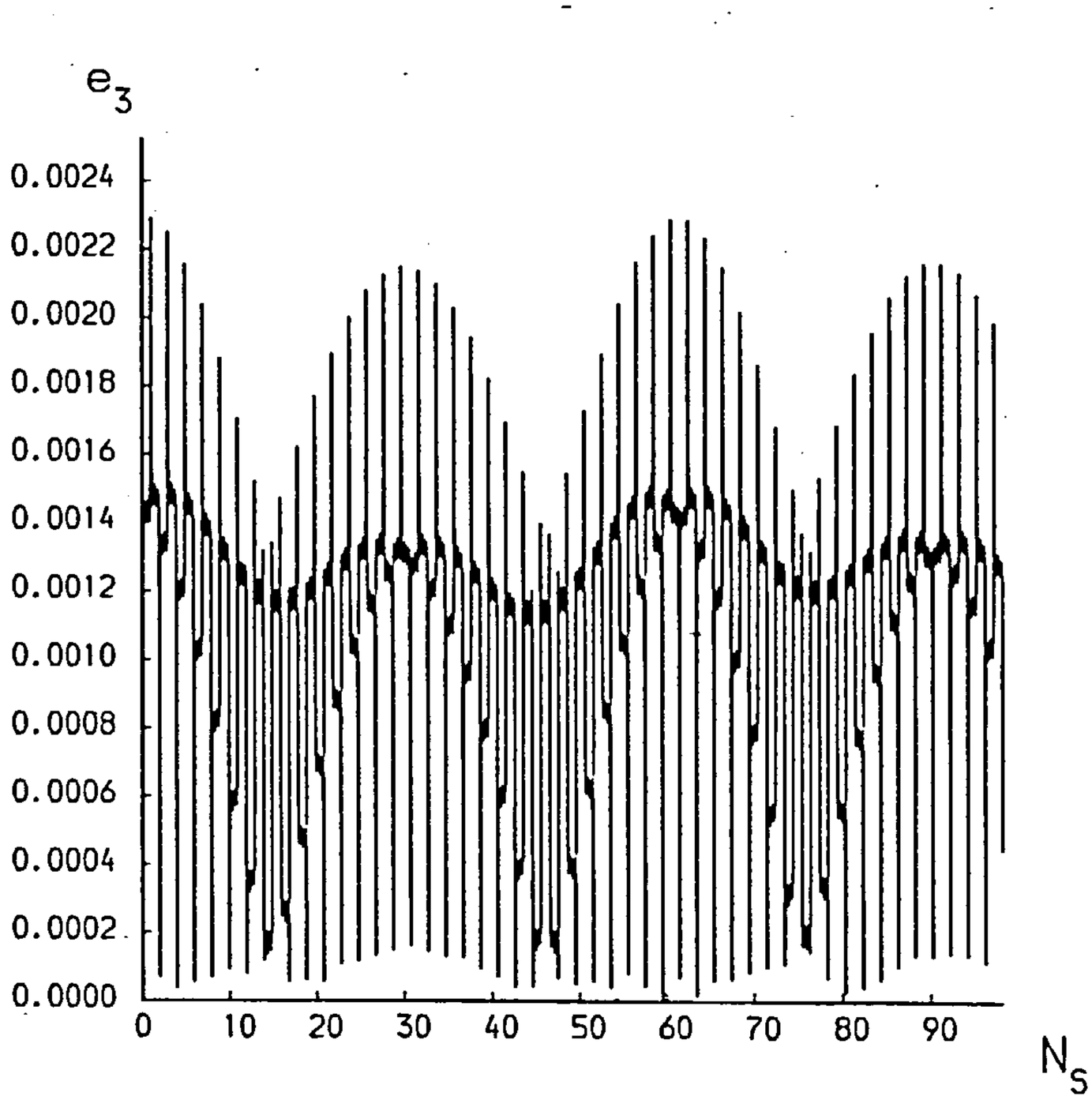
Graph	ε^{23}	ε_{32}	$\alpha(t=0)$	Stable(S)/Unstable (U).
(a)	10^{-4}	10^{-3}	0.926	S
(b)	10^{-4}	10^{-3}	0.942	U
(c)	10^{-4}	10^{-4}	0.972	S
(d)	10^{-4}	10^{-5}	0.97	S
(e)	10^{-4}	10^{-5}	0.98	U
(f)	10^{-6}	10^{-2}	0.9	S
(g)	10^{-6}	10^{-2}	0.948	U
(h)	10^{-6}	10^{-2}	0.955	S(?)


a_2
0.926
0.924
0.922
0.920
0.918
0.916
0.914
0.912
0
10
20
30
40
50
60
70
80
90
N_S

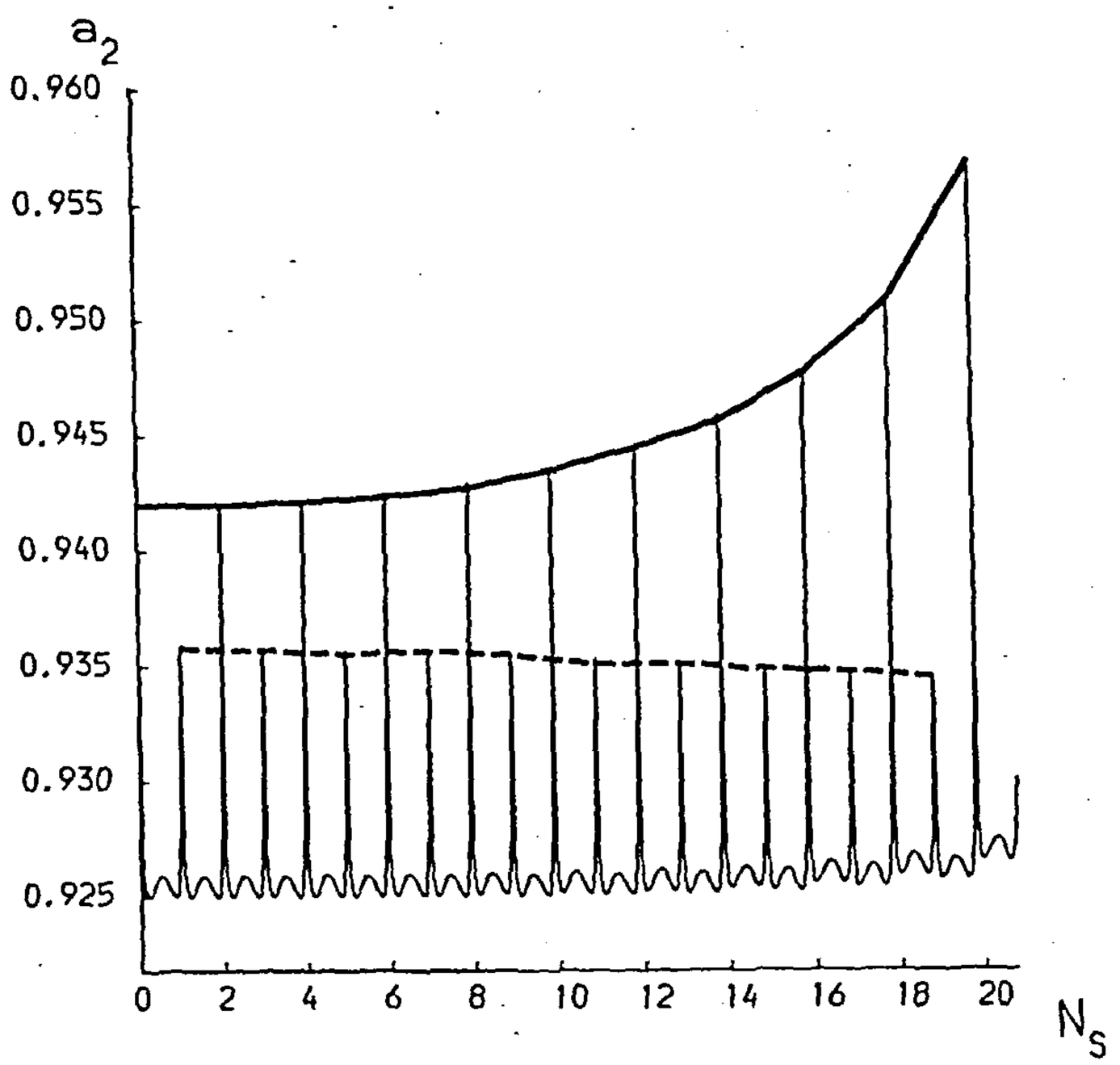

a_3
1.0000
0.9998
0.9996
0.9994
0.9992
0.9990
0.9988
0.9986
0
10
20
30
40
50
60
70
80
90
N_S

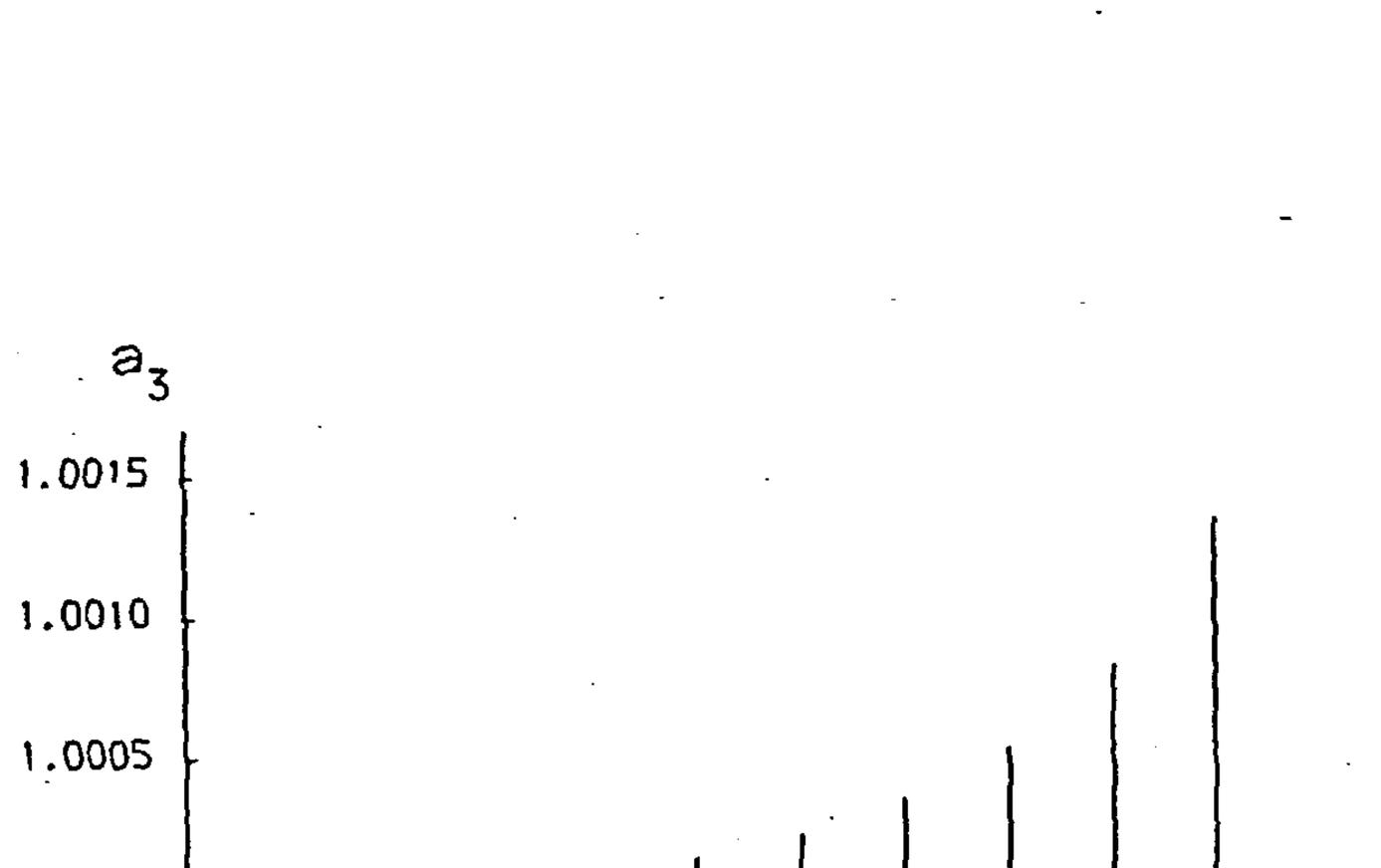
(a)


e_2
0.026
0.024
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
10
20
30
40
50
60
70
80
90
N_S

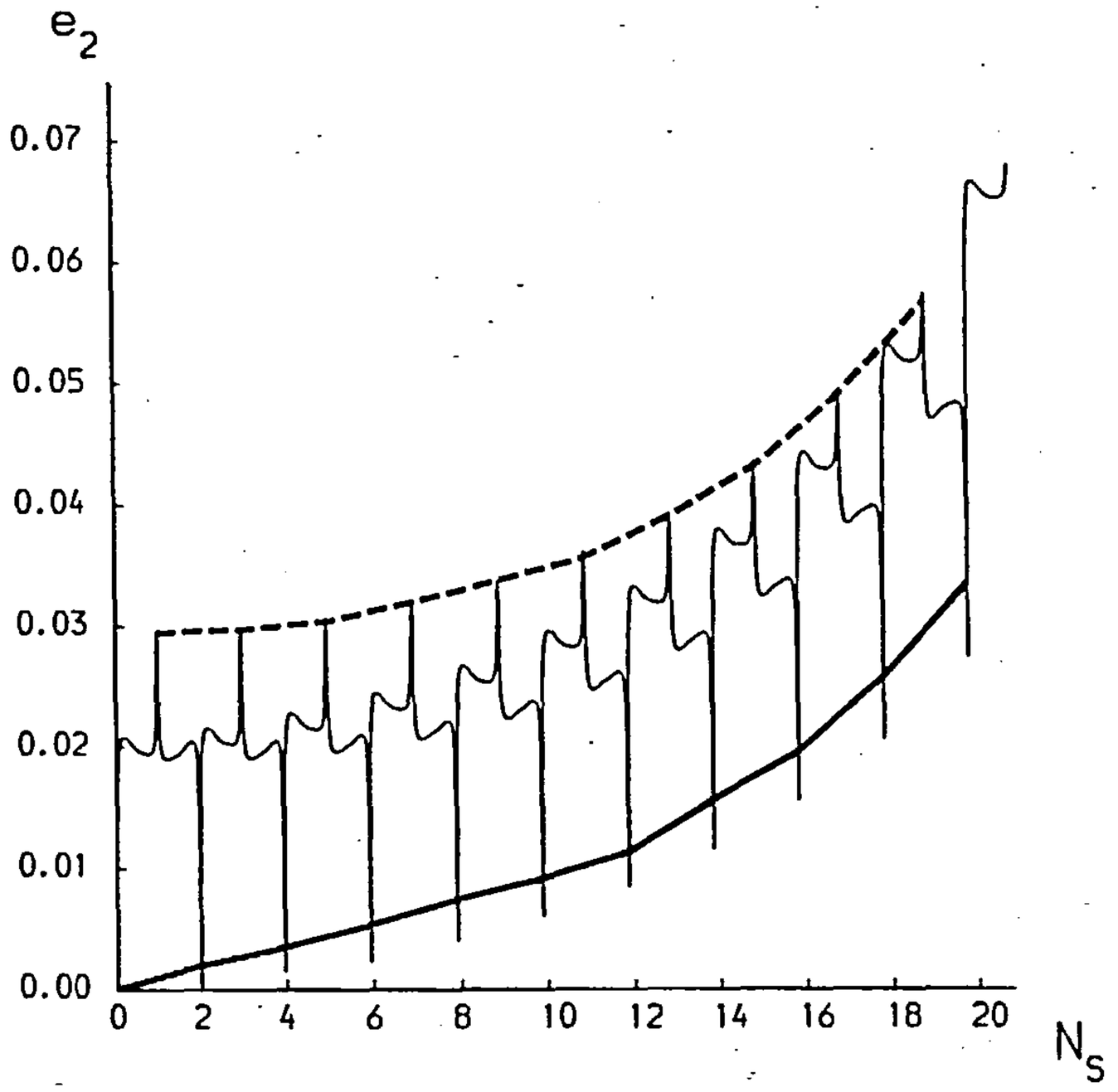

e_3
0.0024
0.0022
0.0020
0.0018
0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
10
20
30
40
50
60
70
80
90
N_S

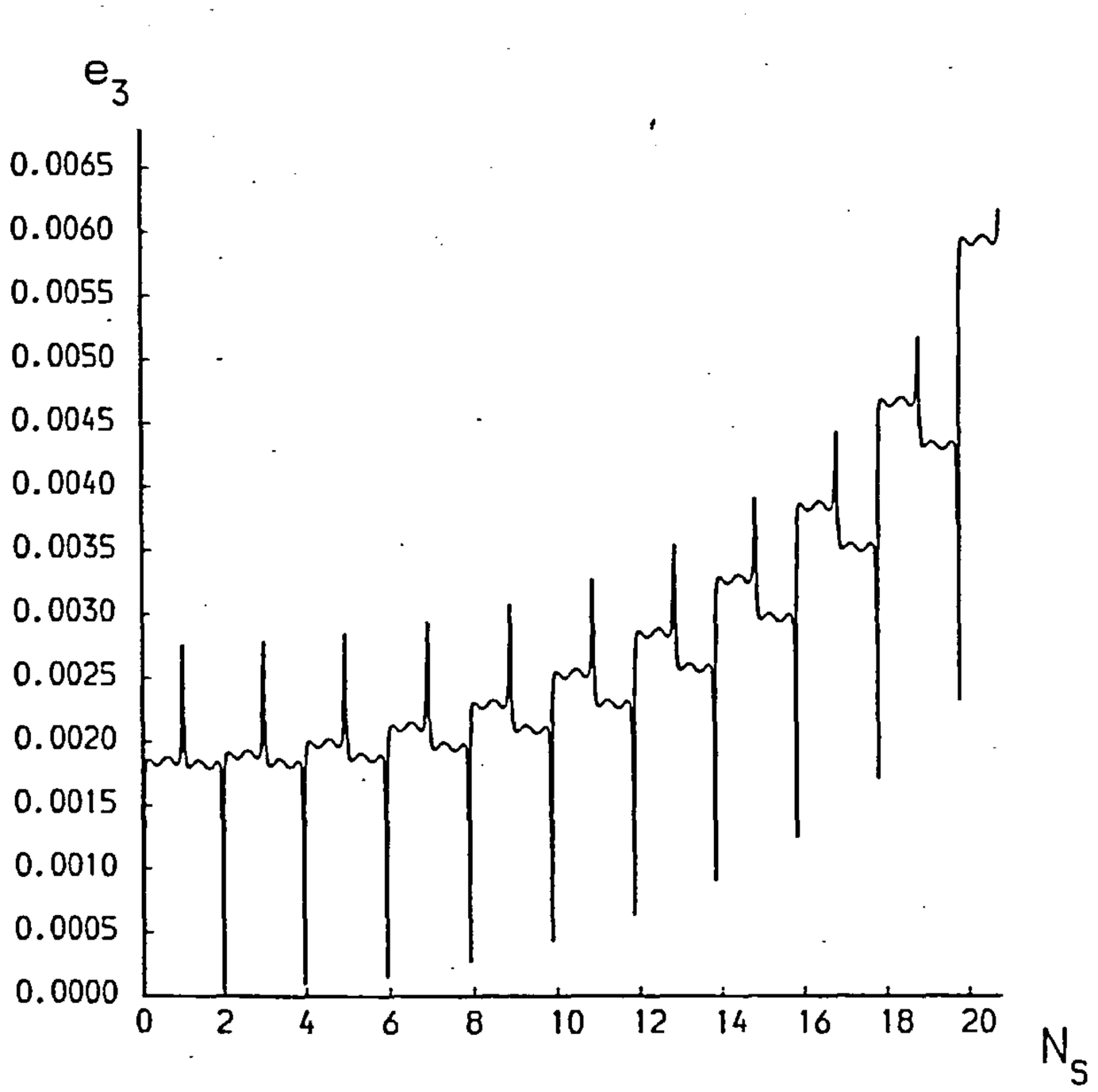
(a)


a_2
0.960
0.955
0.950
0.945
0.940
0.935
0.930
0.925
0
2
4
6
8
10
12
14
16
18
20
N_S

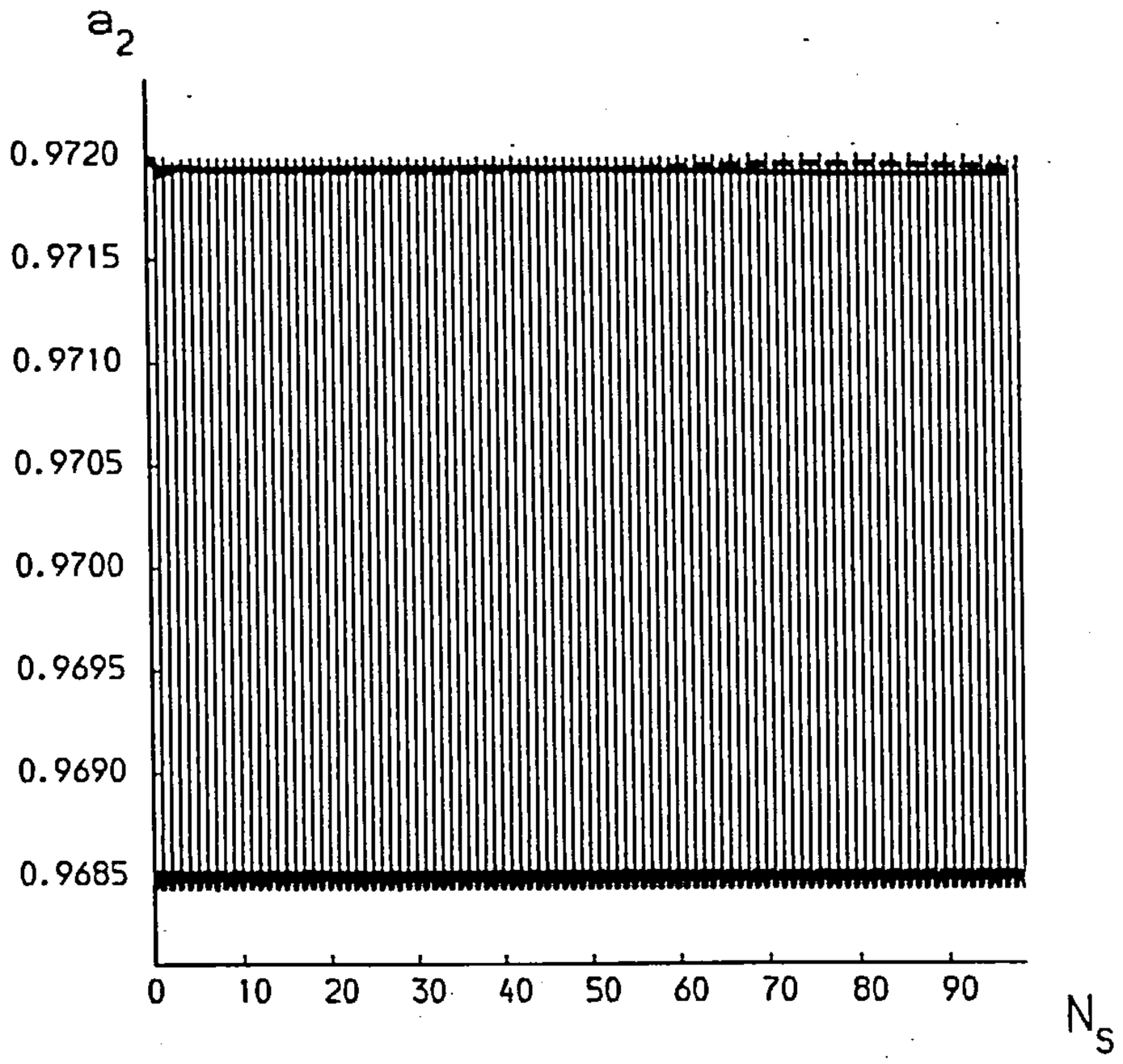

a_3
1.0015
1.0010
1.0005
1.0000
0.9995
0.9990
0.9985
0.9980
0
2
4
6
8
10
12
14
16
18
20
N_S

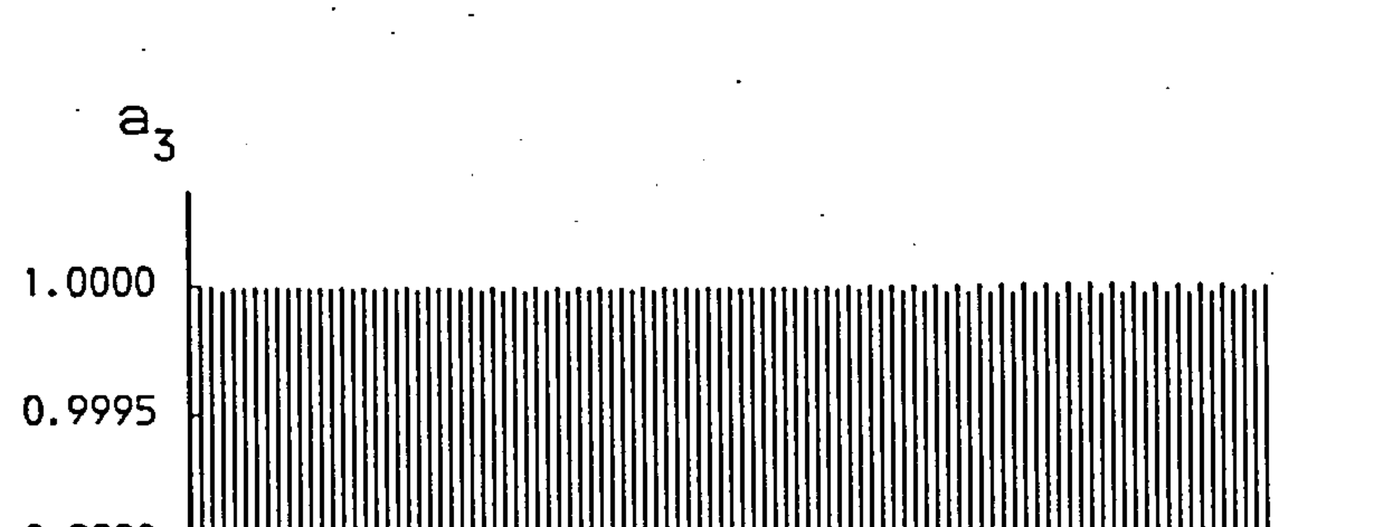
(b)


e_2
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
2
4
6
8
10
12
14
16
18
20
N_S

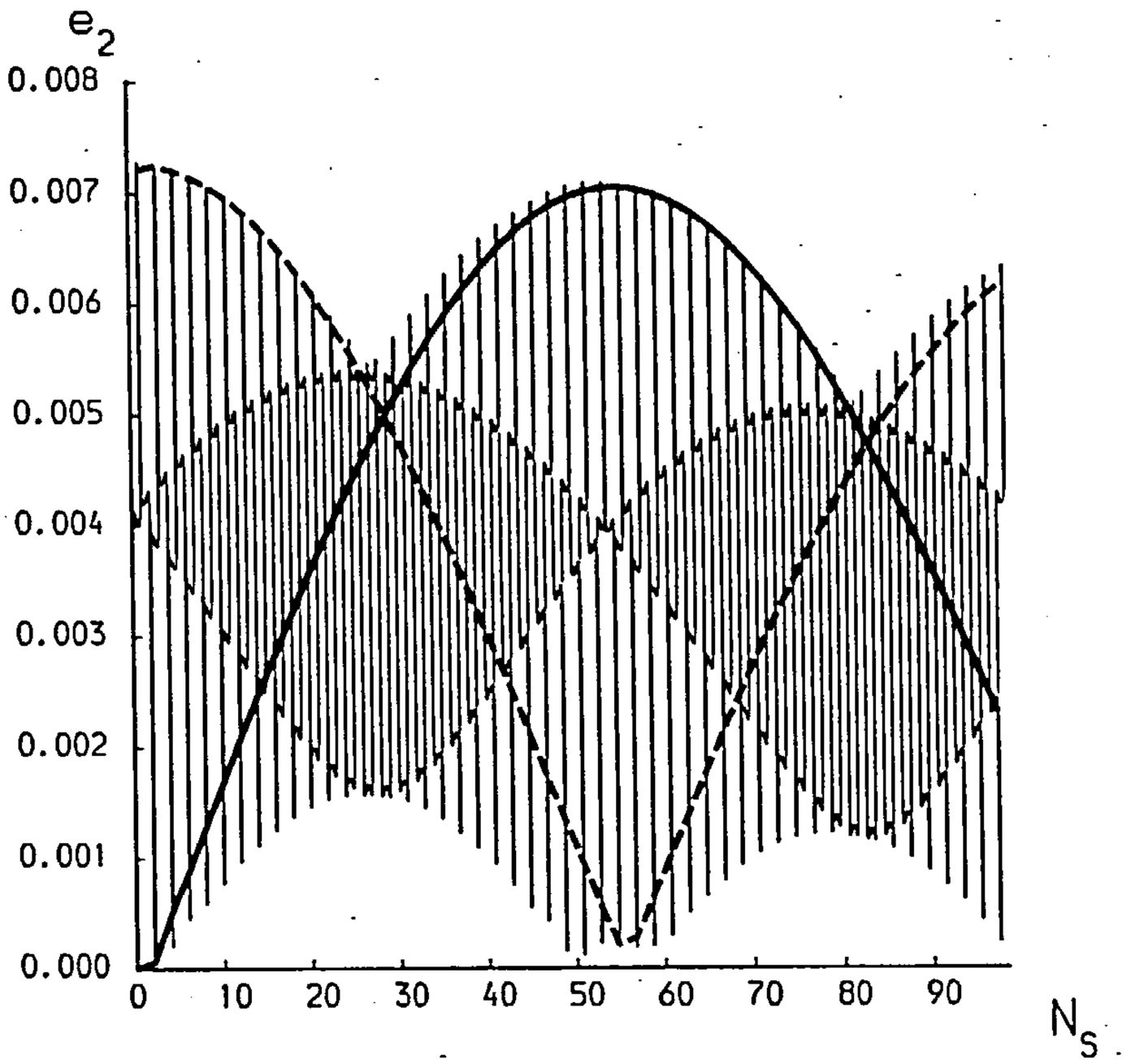

e_3
0.0065
0.0060
0.0055
0.0050
0.0045
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
2
4
6
8
10
12
14
16
18
20
N_S

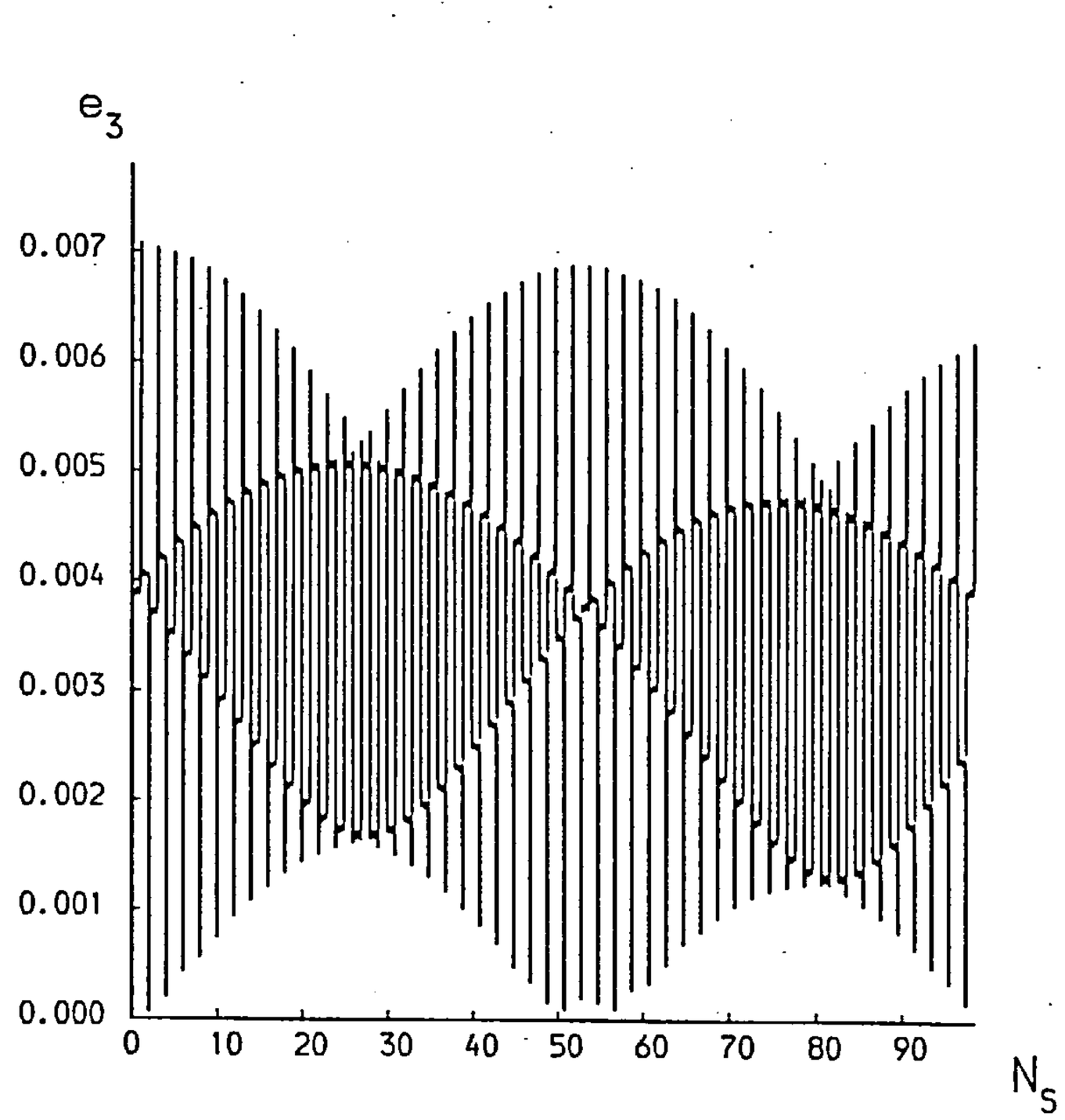
(b)


a_2
0.9720
0.9715
0.9710
0.9705
0.9700
0.9695
0.9690
0.9685
0
10
20
30
40
50
60
70
80
90
N_S

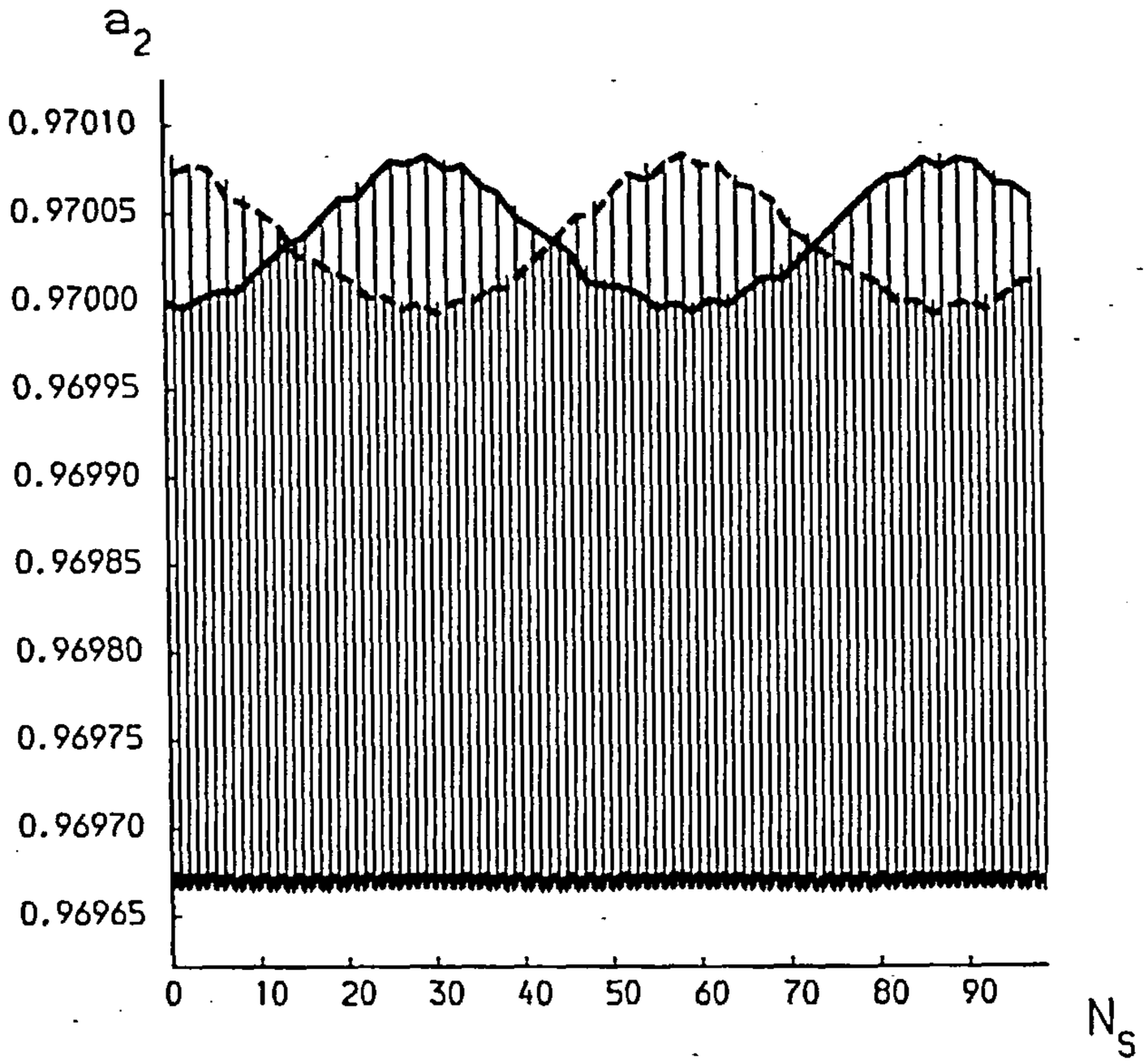

a_3
1.0000
0.9995
0.9990
0.9985
0.9980
0.9975
0.9970
0.9965
0
10
20
30
40
50
60
70
80
90
N_S

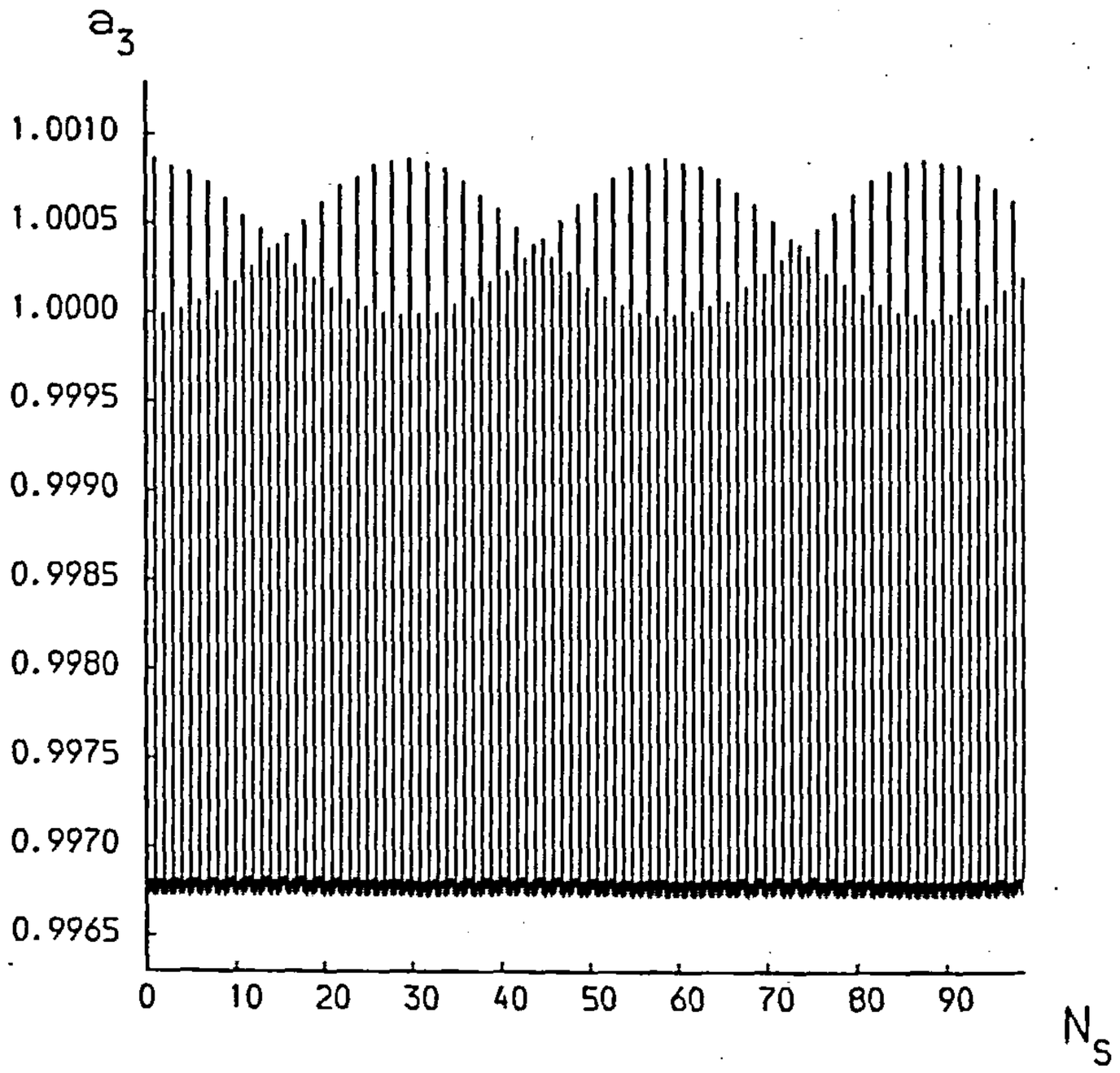
(c)


e_2
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
10
20
30
40
50
60
70
80
90
N_S

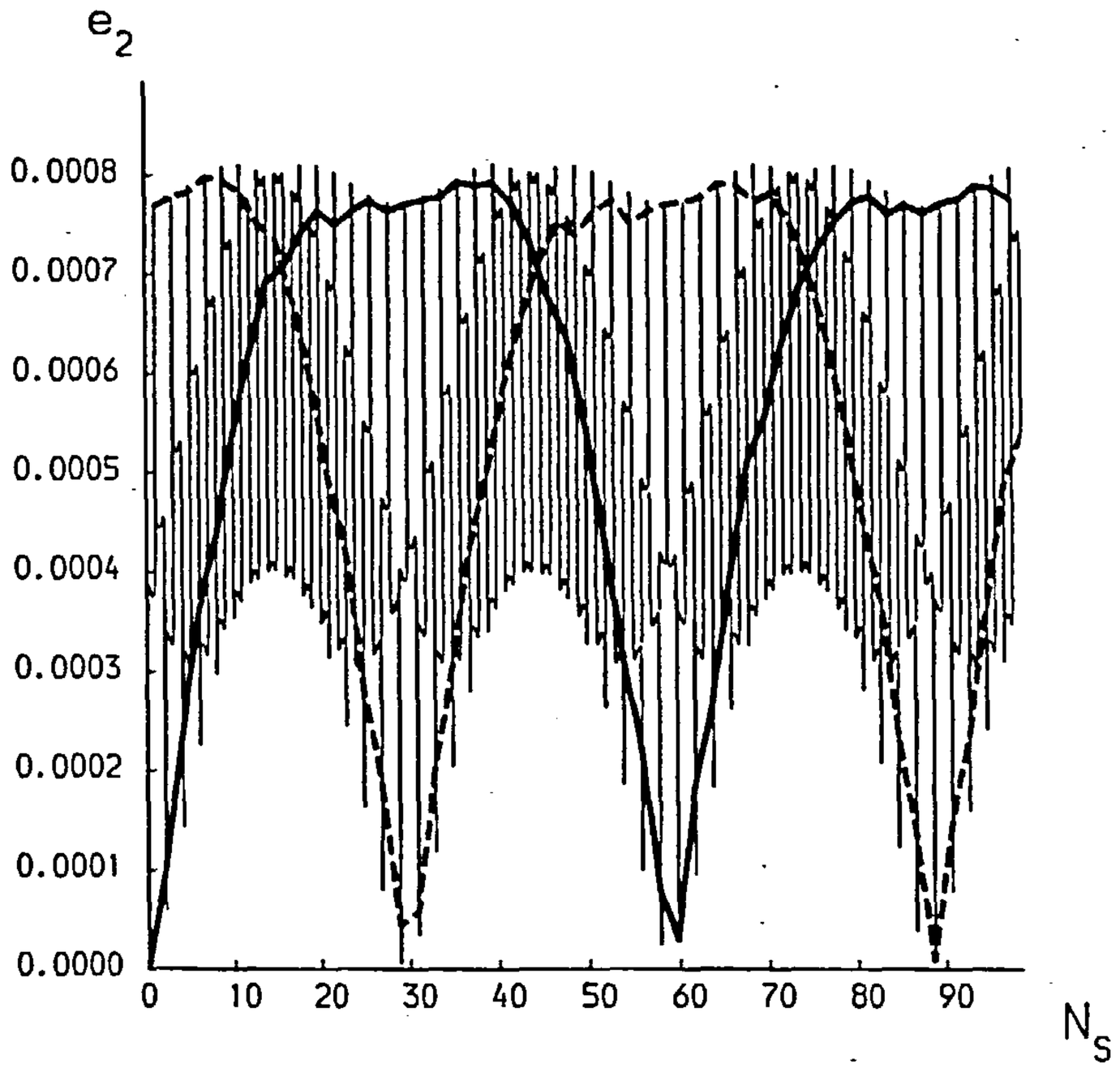

e_3
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
10
20
30
40
50
60
70
80
90
N_S

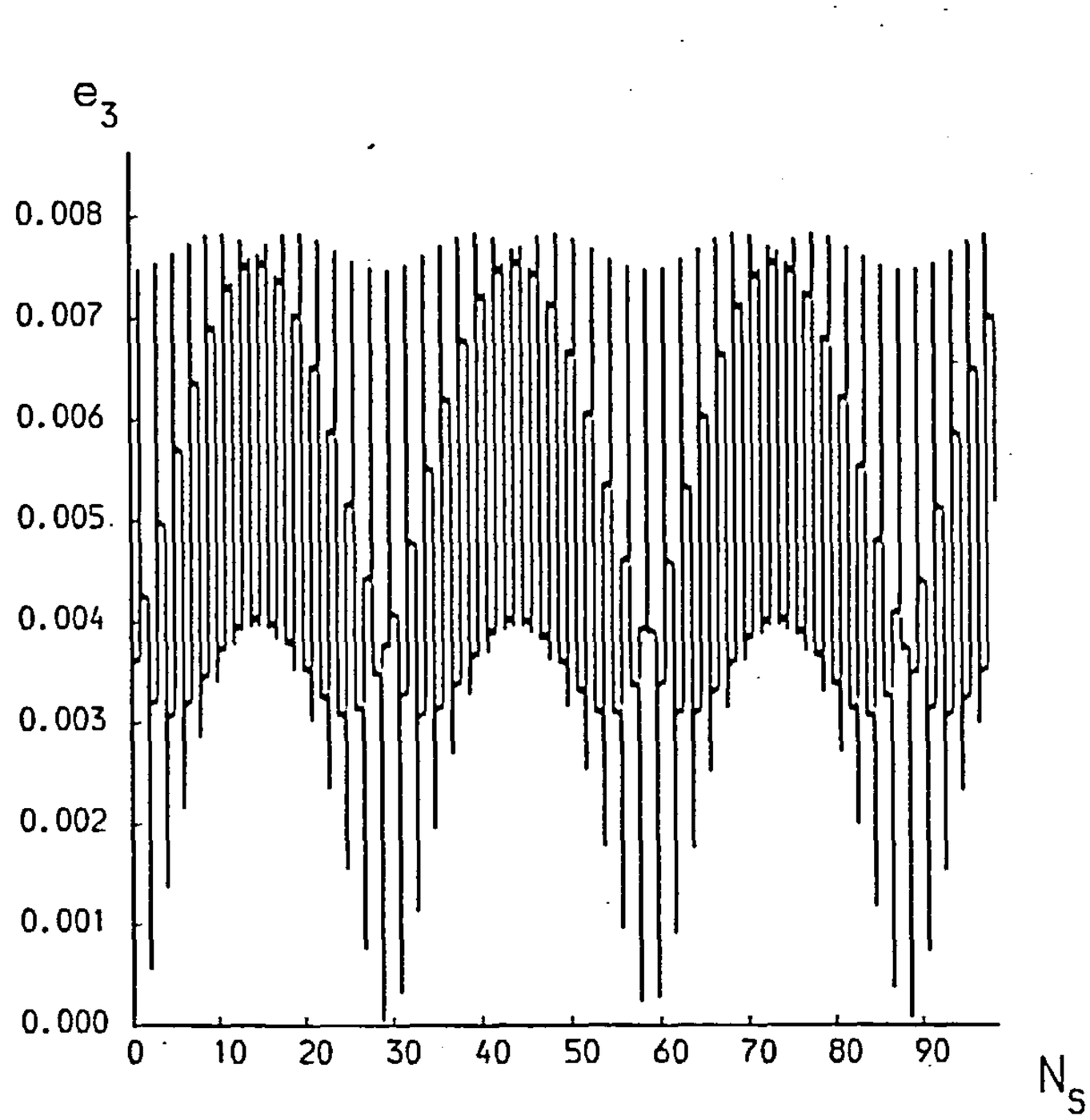
(c)


a_2
0.97010
0.97005
0.97000
0.96995
0.96990
0.96985
0.96980
0.96975
0.96970
0.96965
0
10
20
30
40
50
60
70
80
90
N_S

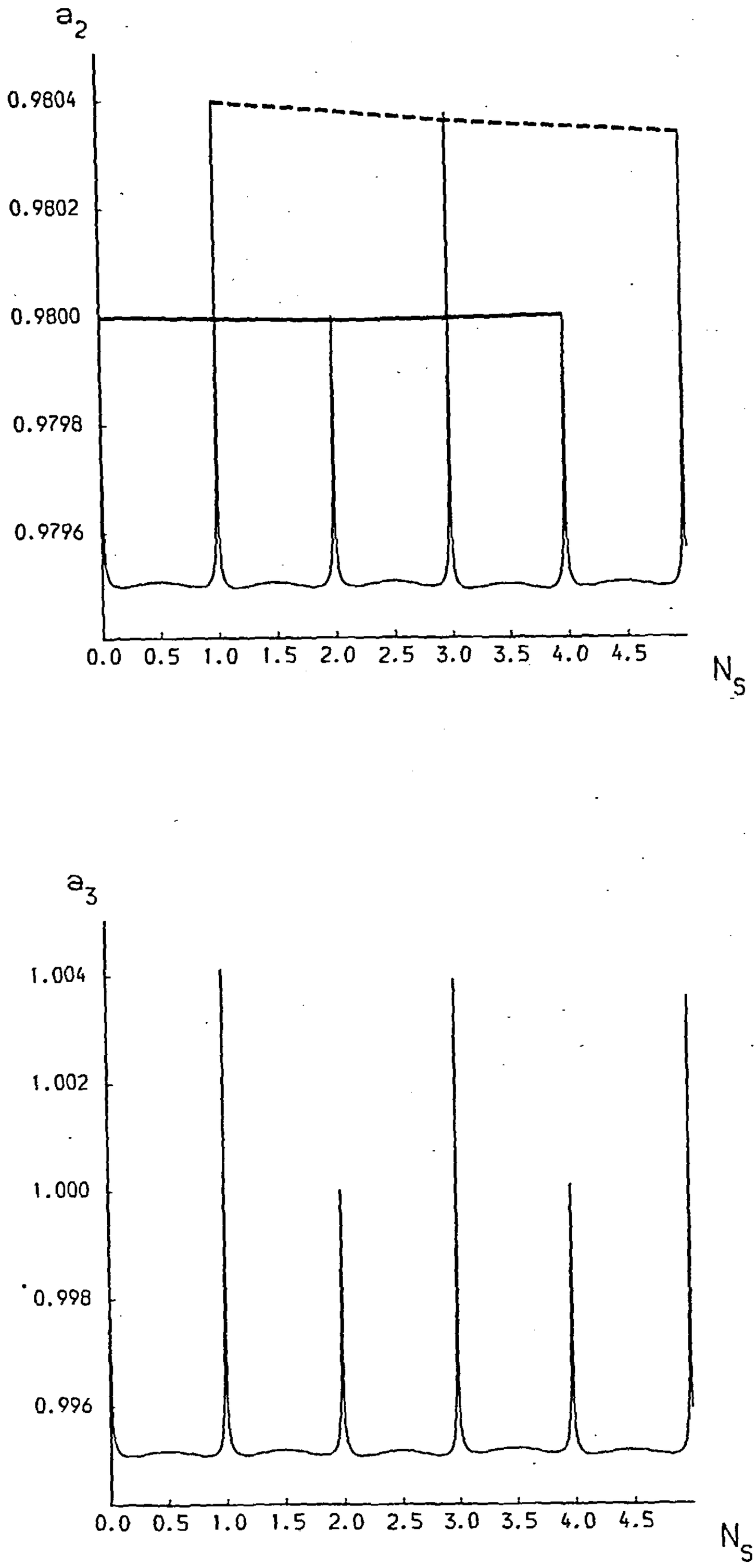

a_3
1.0010
1.0005
1.0000
0.9995
0.9990
0.9985
0.9980
0.9975
0.9970
0.9965
0
10
20
30
40
50
60
70
80
90
N_S

(d)

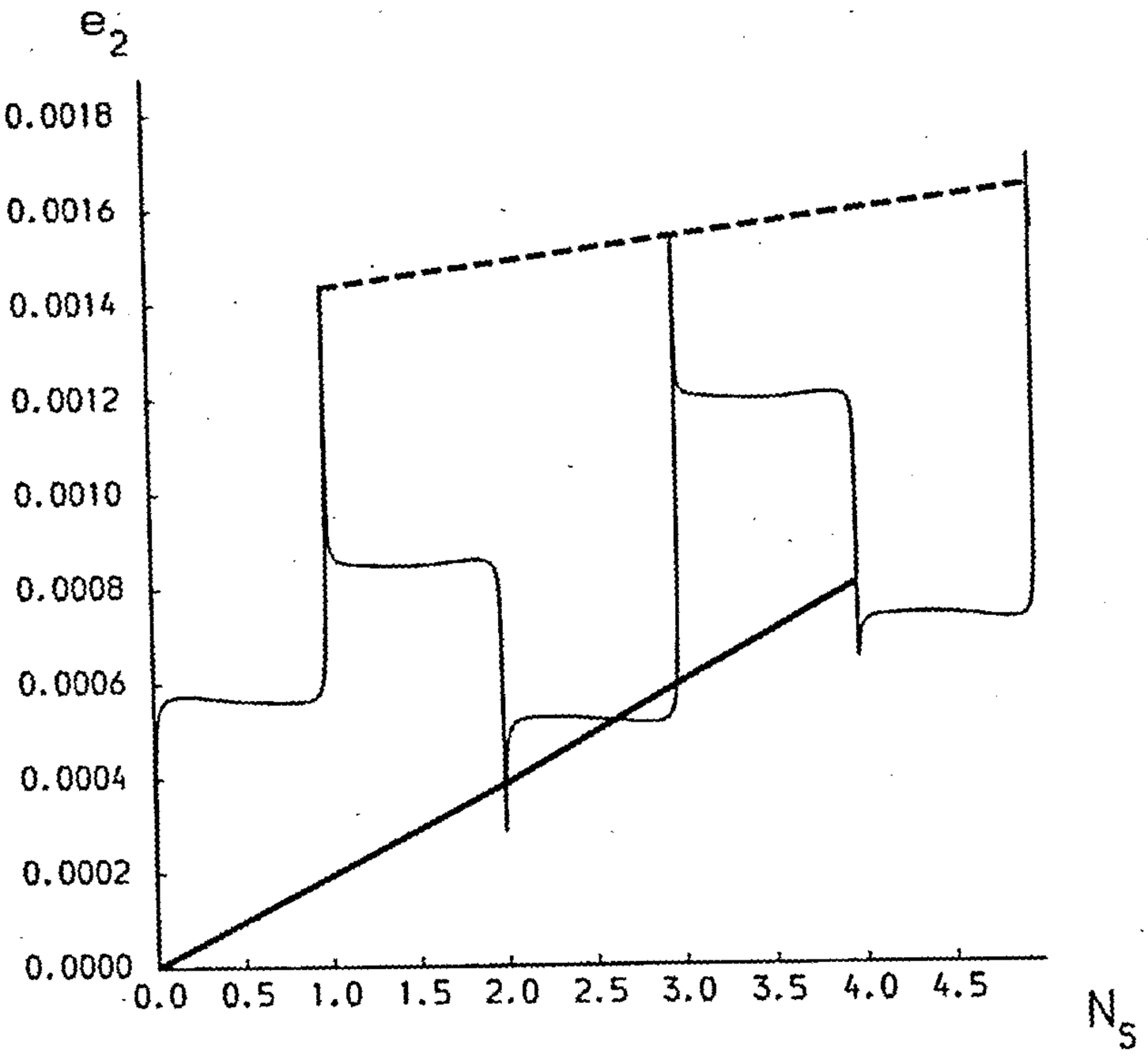

e_2
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0
10
20
30
40
50
60
70
80
90
N_S

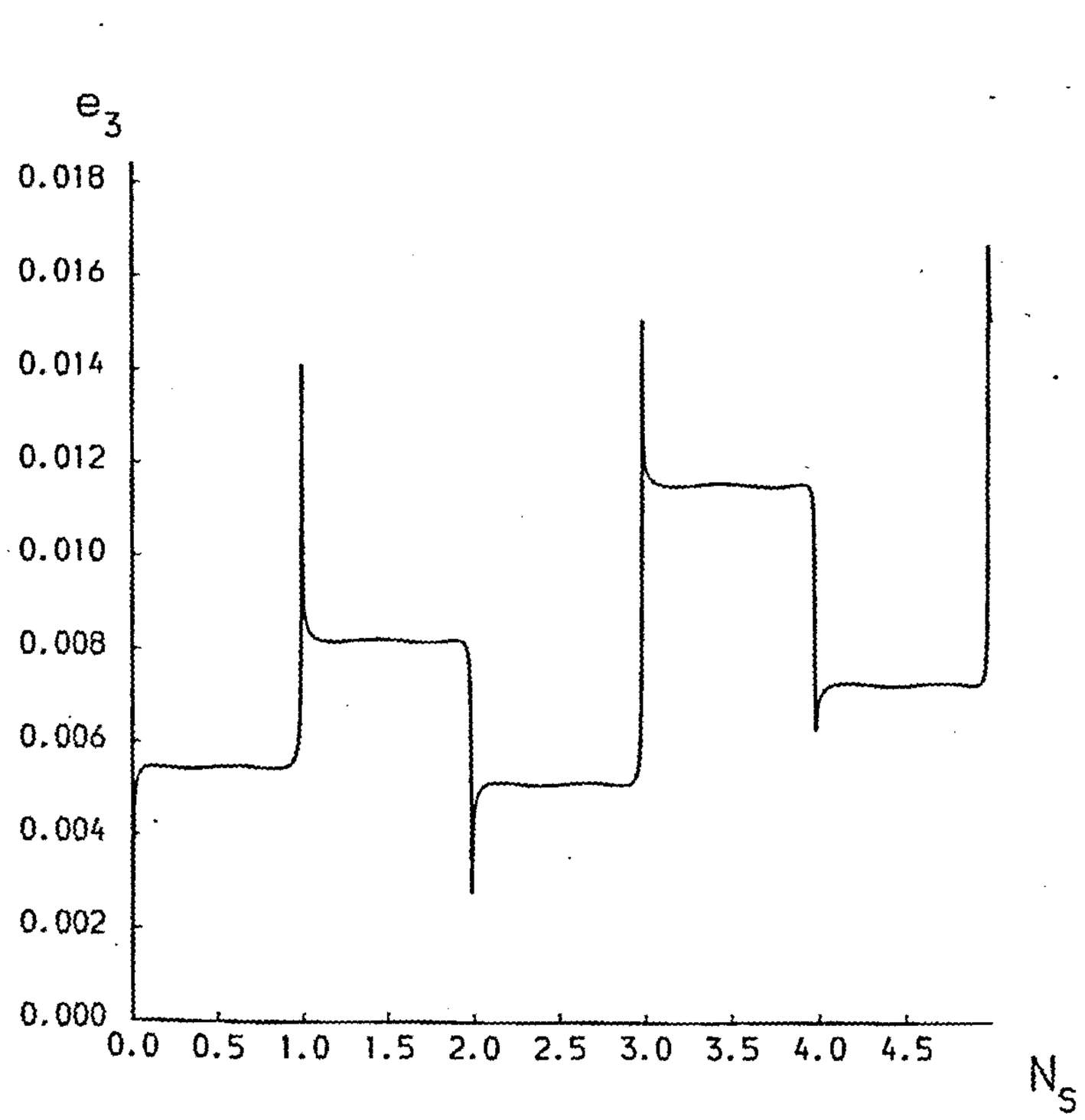

e_3
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
10
20
30
40
50
60
70
80
90
N_S

(d)

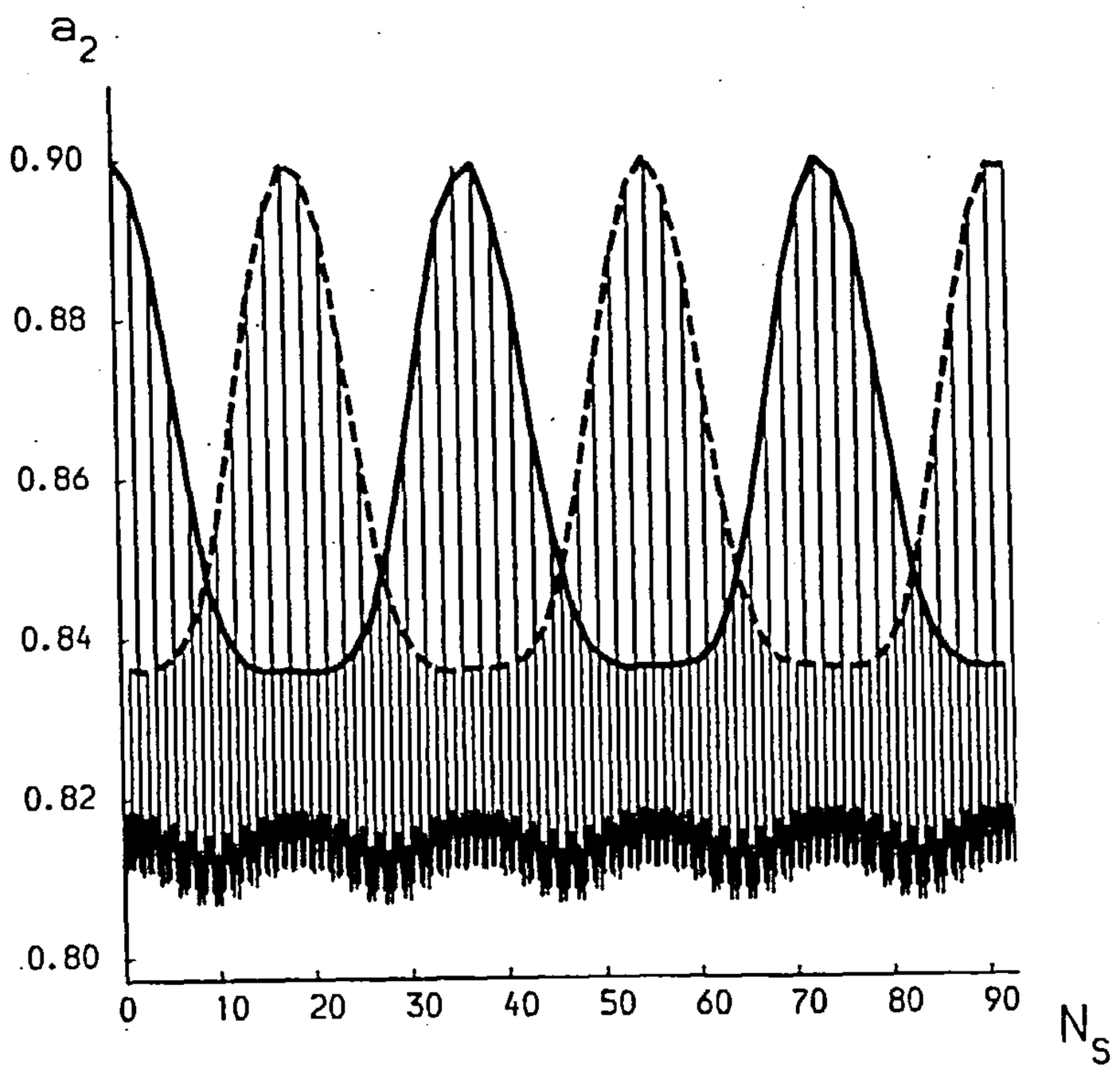

a_2
0.9804
0.9802
0.9800
0.9798
0.9796
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
N_S
a_3
1.004
1.002
1.000
0.998
0.996
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
N_S

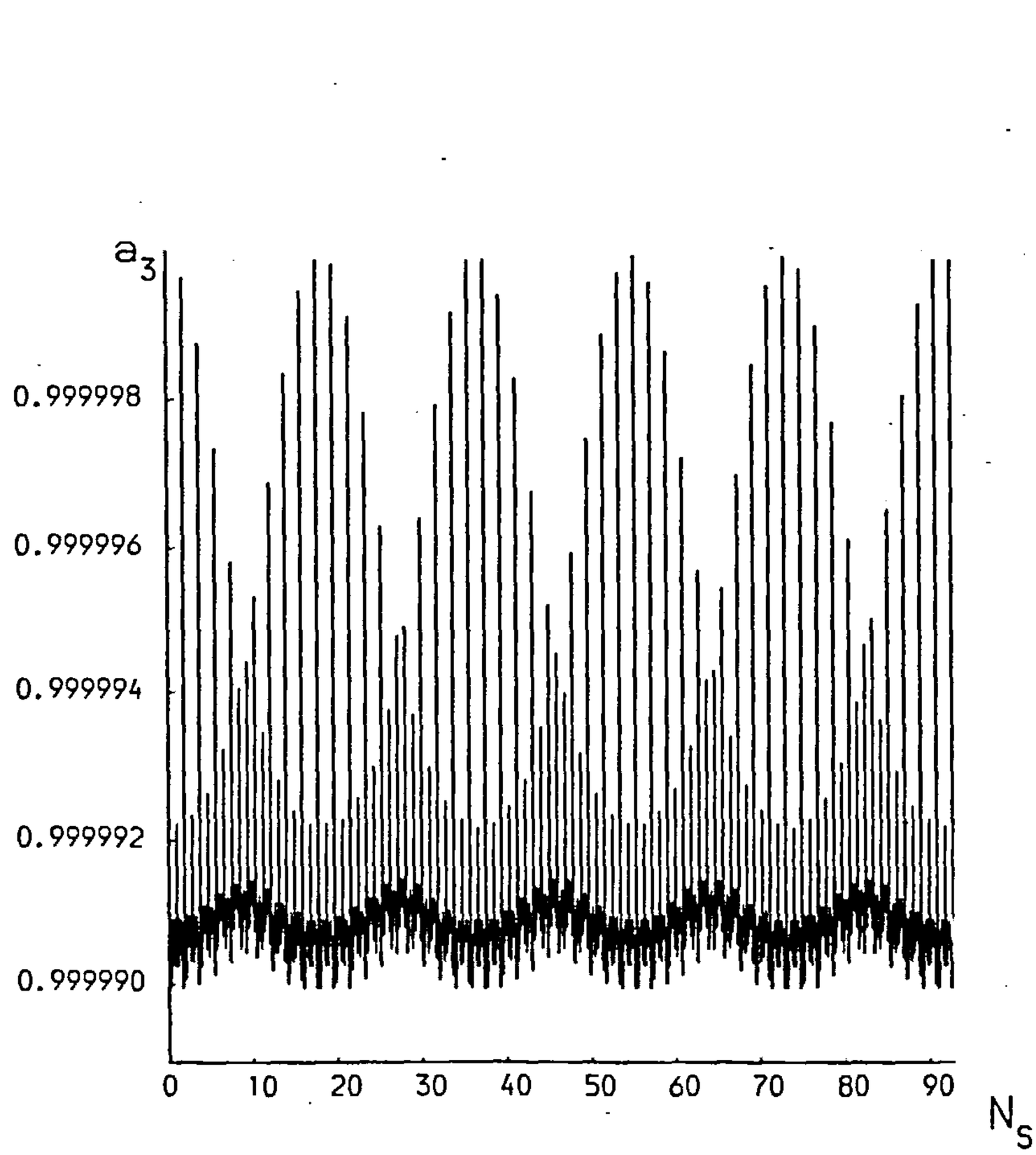
(e)


e_2
0.0018
0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
N_S

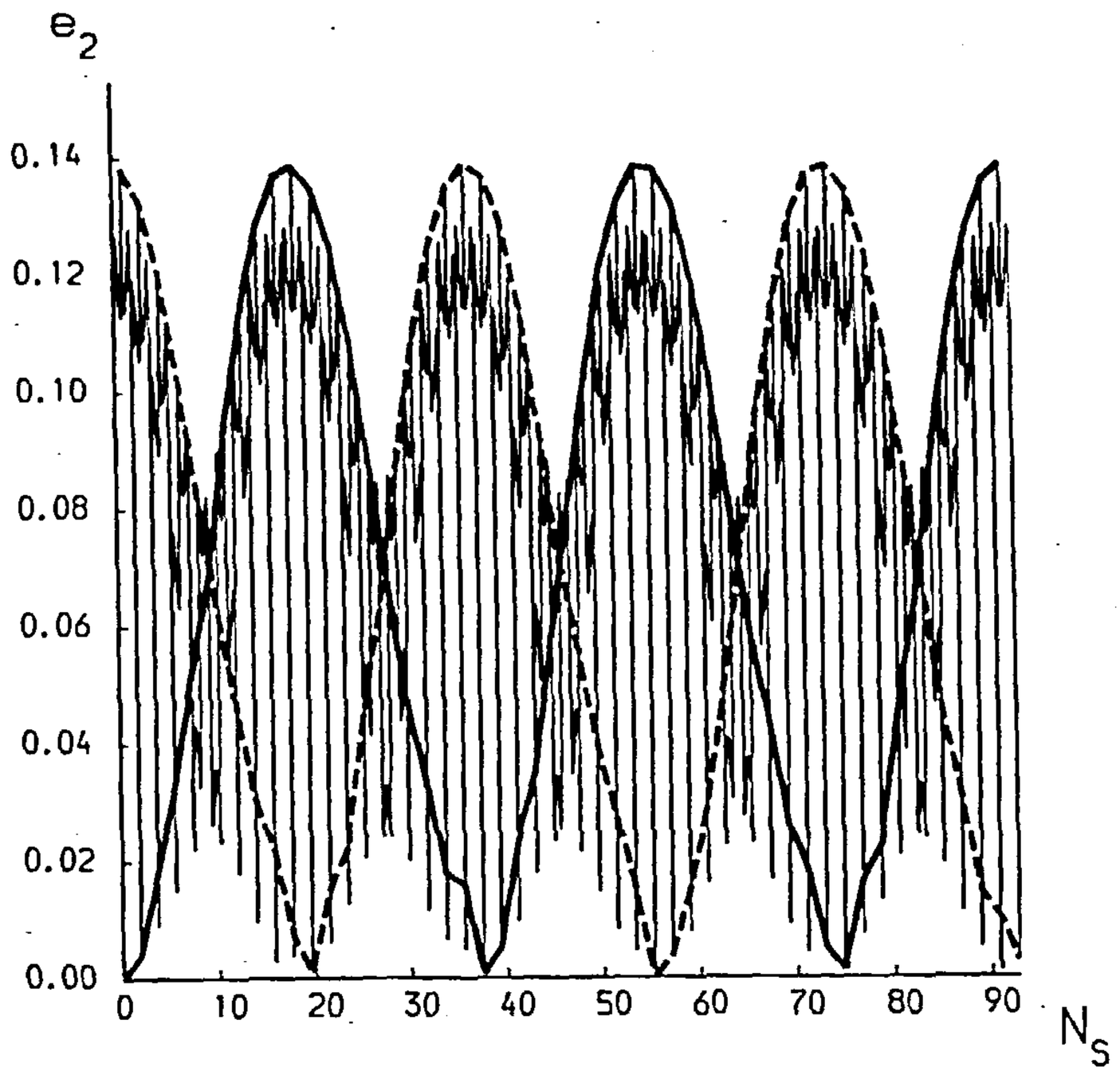

e_3
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
N_S

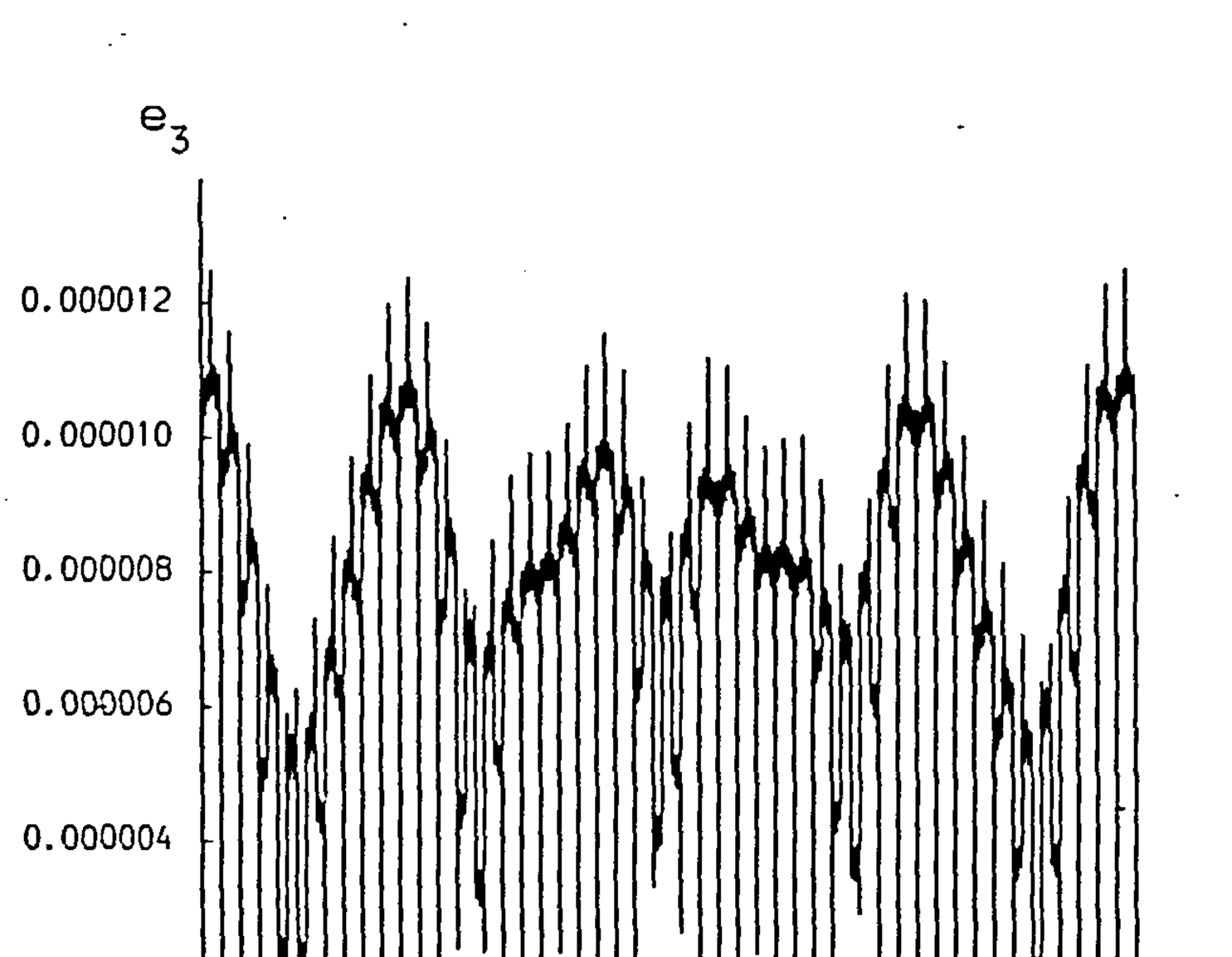
(e)


a_2
0.90
0.88
0.86
0.84
0.82
0.80
0
10
20
30
40
50
60
70
80
90
N_S

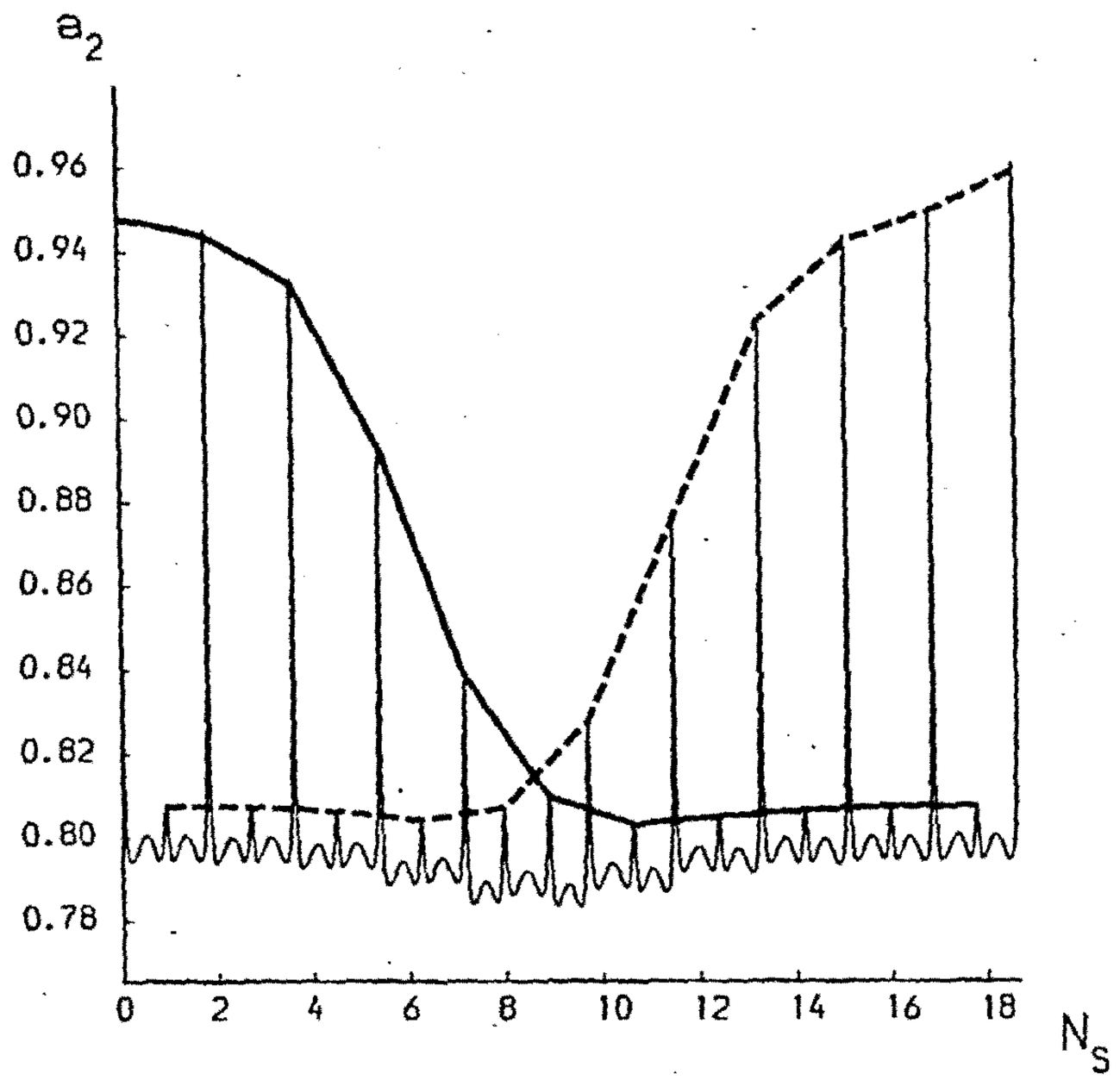

a_3
0.999998
0.999996
0.999994
0.999992
0.999990
0
10
20
30
40
50
60
70
80
90
N_S

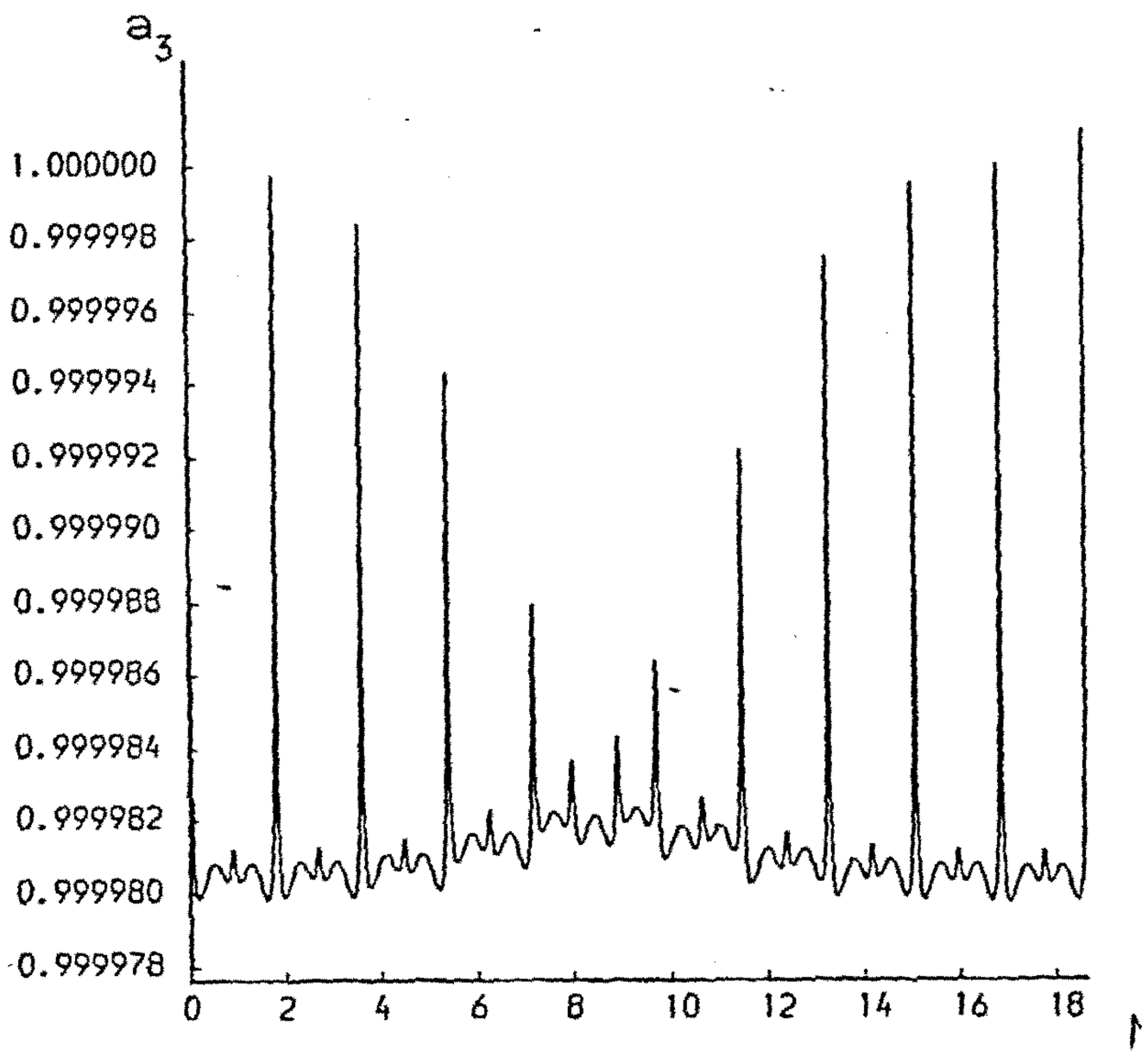
(f)


e_2
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
60
70
80
90
N_S

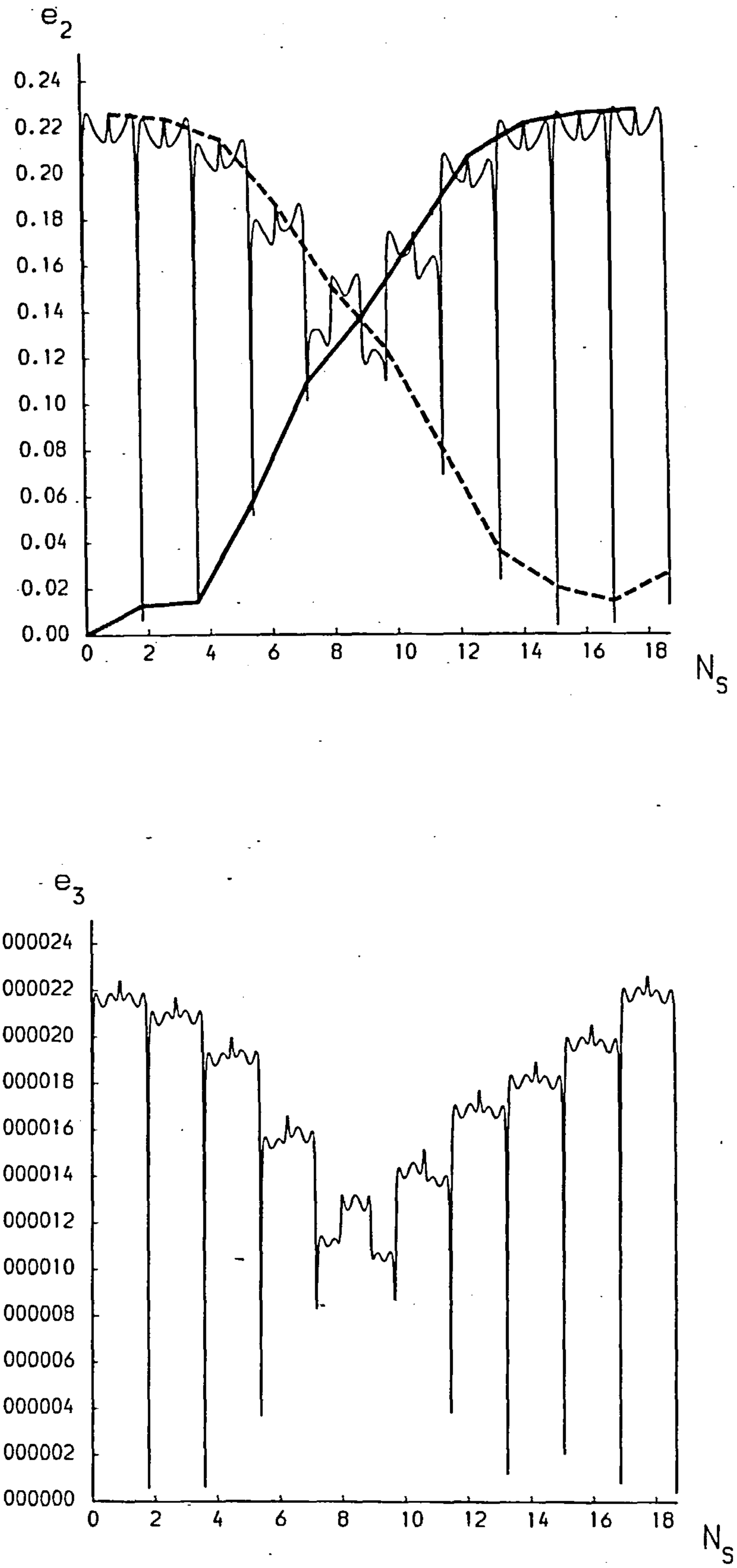

e_3
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
10
20
30
40
50
60
70
80
90
N_S

(f)

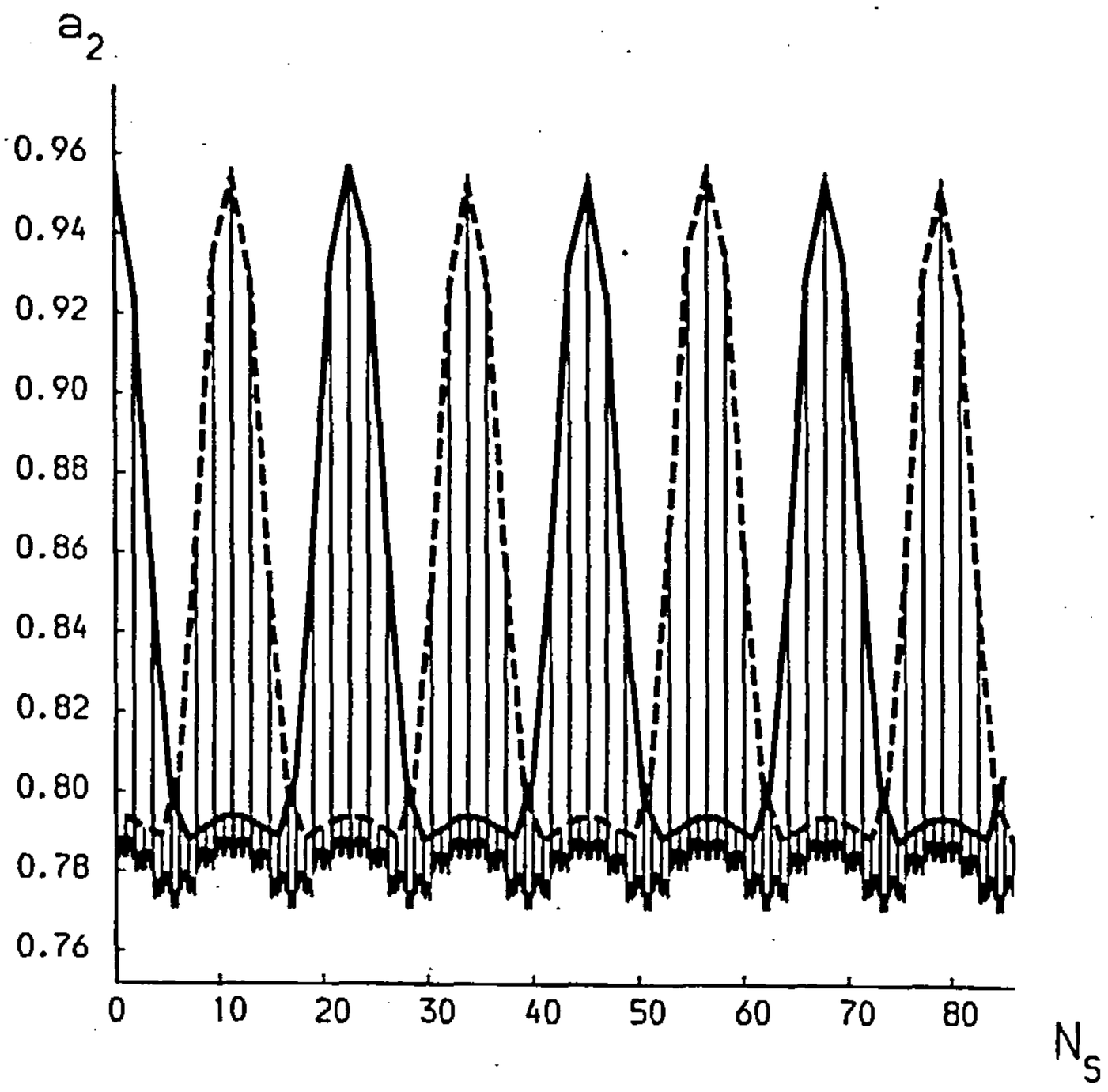

a_2
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0
2
4
6
8
10
12
14
16
18
N_S


a_3
1.000000
0.999998
0.999996
0.999994
0.999992
0.999990
0.999988
0.999986
0.999984
0.999982
0.999980
0.999978
0
2
4
6
8
10
12
14
16
18

(g)

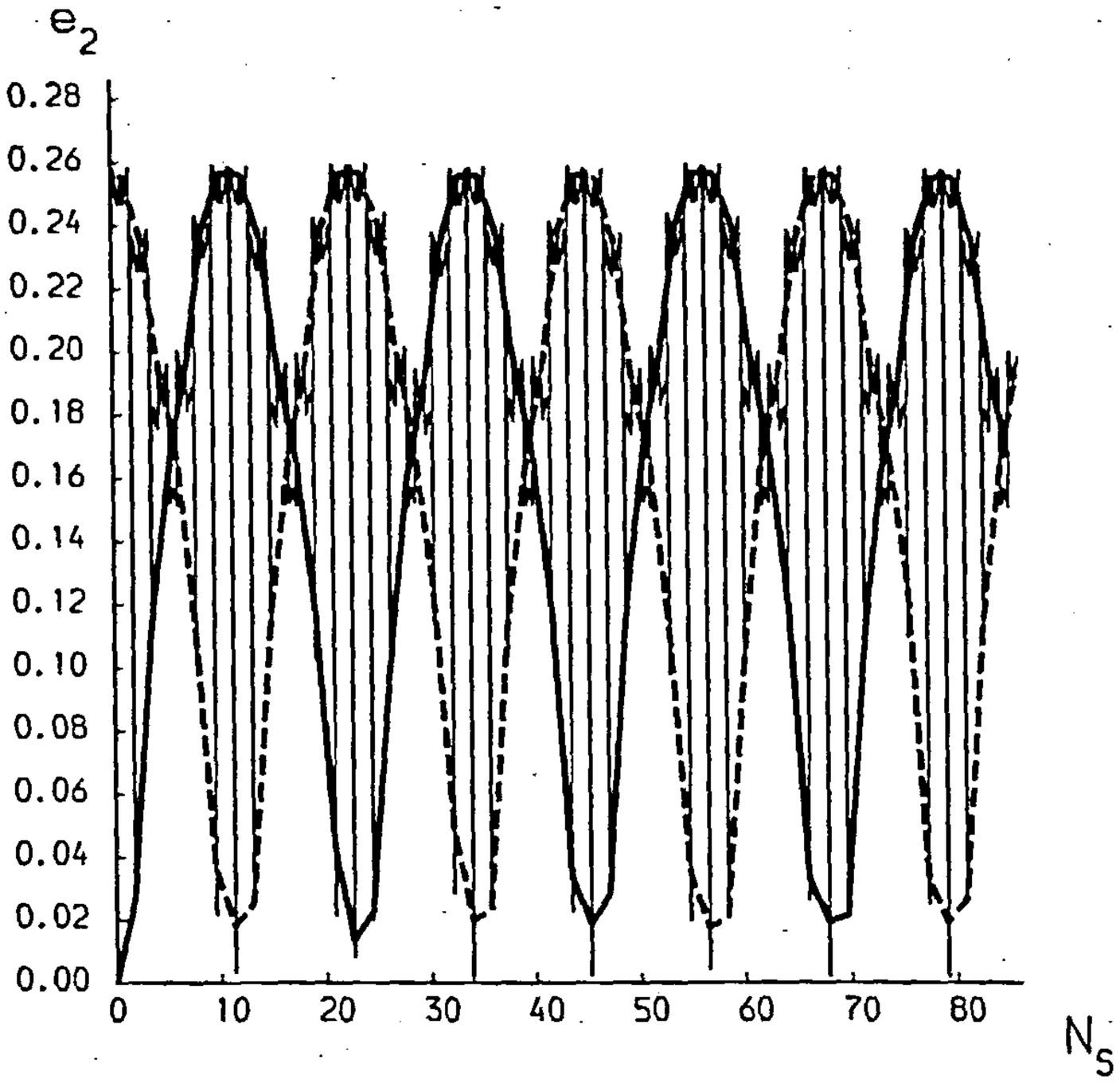

e_2
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
10
12
14
16
18
N_S
e_3
0.000024
0.000022
0.000020
0.000018
0.000016
0.000014
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
2
4
6
8
10
12
14
16
18
N_S

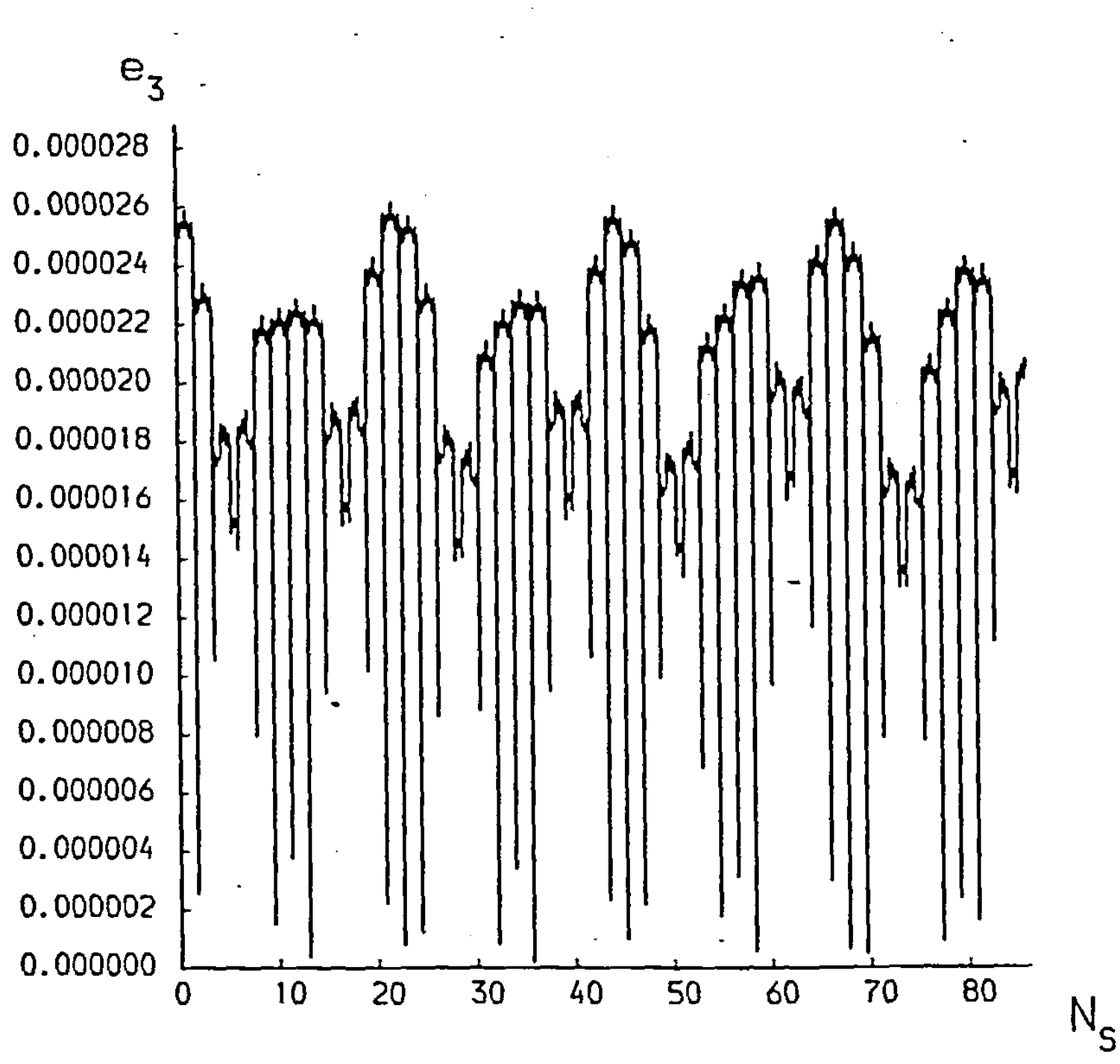
(g)


a_2
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0.76
0
10
20
30
40
50
60
70
80
N_S


a_3
1.000000
0.999998
0.999996
0.999994
0.999992
0.999990
0.999988
0.999986
0.999984
0.999982
0.999980
0.999978
0.999976
0.999974
0
10
20
30
40
50
60
70
80
N_S

(h)


e_2
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
60
70
80
N_S


e_3
0.000028
0.000026
0.000024
0.000022
0.000020
0.000018
0.000016
0.000014
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000
0
10
20
30
40
50
60
70
80
N_S

(h)

4.5

"spike" at conjunction. (Note that the spikes occur at real conjunctions. The time is measured in synodic periods where the synodic period is calculated from the initial conditions. Therefore the spikes will not coincide exactly with the integer values of N_s). For the values of the semi-major axes, the maxima occur near conjunction and the average "running" value is approximately minimum.

For the eccentricities, the behaviour is somewhat different. Both the maxima and minima occur near conjunction with the running value(s) being somewhere between the two extremes. Quite often the running value after a minimum differs from the running value after a maximum. These differences may vary over many synodic periods and this gives rise to some of the very beautiful eccentricity graphs produced by the stable systems (a), (c), (d), (f), (h).

There is one noticeable difference between the behaviours of e_2 and e_3. Between conjunctions, e_2 has two turning points and e_3 has four. None of these turning points occur at opposition. There is no corresponding difference between a_2 and a_3. There is no immediately obvious physical explanation for this behaviour.

The magnitude of the perturbation on the osculating elements in each synodic period varies with ε^{23} and ε_{32}. This is not surprising since ε^{23} and ε_{32} are measures of the disturbance on the outer and inner binary orbits respectively. Defining $\delta a = (a_{max} - a_{min})/a_{min}$ and $\delta e = e_{max} - e_{min}$ for inner and outer orbit, we find that after one synodic period

$$\delta a_2 = p\ \varepsilon_{32}$$

$$\delta a_3 = p\ \varepsilon^{23}$$

$$\delta e_2 = q\ \varepsilon_{32}$$

$$\delta e_3 = q\ \varepsilon^{23}$$

4.5

where p and q vary with ε^{23} , ε_{32}, α but typically have values between 10 and 100. Both p and q increase as α increases. For stable systems, δa_2, δa_3, δe_2, δe_3 change in a periodic fashion only as do the running averages (Figure 4.3(a)). For an unstable system however, the average and/or δa, δe may change in an irreversible manner as the orbit is disturbed. These disturbances will be compounded and lead to a break up of the hierarchy.

Figure 4.3(b) is a good example of this. The peak values of a_2 and a_3, which are attained every even conjunction, rise steadily until crossover is reached. Note also that the running values of a_2 between conjunctions increases, while the running value of a_3 decreases. Combining this with the steadily rising eccentricities, it is not difficult to see that such behaviour inevitably leads to a crossover of the two orbits. The apocentre distance of the inner will increase, while the pericentre distance of the outer will decrease.

Of two unstable systems (with equal epsilons) the one with the lower initial value of α will survive for longer before exhibiting crossover. This is because the lower the value of α to begin with, the greater the initial separation of m_2 and m_3. The system can therefore suffer more of an increase in either a_2 or e_2 before crossover occurs. Since the magnitude of the changes does not vary too much with α , the time to achieve crossover is longer.

This is connected with the question of whether we are able to fit a curve to the (α, N_s) data or not, depending on ε^{23} being greater than, equal to, or less than ε_{32}. Most of the systems that we are investigating have high α values (>0.9). Therefore since

$$\frac{\varepsilon^{23}}{\varepsilon_{32}} = \frac{\mu(1-\mu)}{\mu_3\ \alpha}$$

$$\sim \frac{\mu}{\mu_3} \qquad \text{for} \quad \mu << 1\ , \quad \alpha \simeq 1$$

$$= \frac{m_2}{m_3}\ ,$$

The relative sizes of the epsilons reflect the relative sizes of m_2 and m_3 to a close degree.

For $\varepsilon^{23} < \varepsilon_{32}$, $m_2 < m_3$, (Figure 4.3(b)) and the perturbation of m_3 on the inner binary orbit is greater than that of (m_1, m_2) on the outer binary orbit. In this case, by a combination of small increments in a_2 and e_2, m_2 approaches the m_3 orbit (more so than m_3 approaching m_2), until crossover is reached. The time till crossover depends on the initial separation of m_2 and m_3.

For $\varepsilon^{23} = \varepsilon_{32}$, $m_2 \simeq m_3$, and each body has an equal effect on the other. It seems in this case that either m_2 and m_3 are sufficiently well spaced never to interact or else the effect they have on each other is so strong as to instantly disrupt each other's orbit. Unlike the previous case when m_3 could affect m_2's orbit without any repercussions, m_2 is now big enough to "fight back" and the result is mutual instability of each orbit in a very short time.

When $\varepsilon^{23} > \varepsilon_{32}$, $m_2 > m_3$ instability occurs more rapidly in general compared to the case when $\varepsilon^{23} < \varepsilon_{32}$. Unlike the case when $\varepsilon^{23} = \varepsilon_{32}$, however, there are some systems that survive for a few synodic periods before crossover. Figure 4.3(e) shows one such system. In this case, the running values of a_2 and a_3 seem to be fairly constant. Indeed, the peak values seem to decrease. The eccentricities both rise however and it is this that leads to a crossover. The influence of m_2 on m_3 is much greater than that of m_3 on m_2.

4.5

If we compare the (α, N_s) plots of Figures 4.1 for

(a) $\varepsilon^{23} = 10^{-4}$, $\varepsilon_{32} = 10^{-5}$, and

(b) $\varepsilon^{23} = 10^{-5}$, $\varepsilon_{32} = 10^{-4}$

we find that the value of α_o for (a) is greater than that for (b). However for equal $\alpha > \alpha_o$, the lifetimes for (a) are shorter than for (b). The cut off between stable and unstable is much sharper for (a). It does seem therefore that the outer orbit is more sensitive to a strong perturbation than the inner orbit.

The final point to mention concerning unstable systems is to stress again that the maximum values of the elements occur around conjunction, giving the best opportunity for a crossover of orbits. This is the reason for the "stepping" in the (α, N_s) data shown in Figures 4.1.

Turning now to the behaviour of the stable systems, we see that for constant ε^{23} and ε_{32}, the running value of a_2 for a stable system is lower than that for an unstable system. The running value of a_3 is higher, and the values of e_2 and e_3 are lower. The inner and outer orbits are always more widely spaced, preventing crossover and enhancing stability. We are now able to study the elements over much longer times. When we do, many interesting features become apparent (Figures 4.3(a), (c), (d)). We have already noticed that the elements a,e exhibit fluctuations with a period of two synodic periods, there being peak values at conjunctions.

There was a substantial difference between the $(a_2, a_3, e_2, e_3)o$ curves and the $(a_2, a_3, e_2, e_3)e$ curves. Over many synodic periods, a long period trend becomes apparent. Both odd and even curves oscillate about a common mean value but with opposite phase. The

period of oscillation is common to a_2, a_3, e_2 and e_3. Just after the start of each integration,

$$(e)_o > (e)_e \quad , \quad \forall \varepsilon^{23}, \varepsilon_{32}$$

$$(a)_o < (a)_e \qquad \varepsilon^{23} < \varepsilon_{32}$$

$$(a)_o > (a)_e \qquad \varepsilon^{23} > \varepsilon_{32}$$

$$(a)_o = (a)_e \qquad \varepsilon^{23} = \varepsilon_{32} \ .$$

for both inner and outer binary orbits. In addition, there seems to be no oscillation of $(a)_o$ and $(a)_e$ when $\varepsilon^{23} = \varepsilon_{32}$. Clearly the phase and amplitude of this long period oscillation is linked to the size and mutual difference of ε^{23} and ε_{32}.

The following is a physical interpretation of some of these results. Firstly, define T_2 to be the sidereal period of m_2 about m_1 and T_3 to be the sidereal period of m_3 about M_2. $n_i = 2\pi/T_i$ are the corresponding mean motions. The synodic period for a <u>retrograde</u> system S is defined by

$$\frac{1}{S} = \frac{1}{T_2} + \frac{1}{T_3} \qquad (7)$$

Note that $S < T_2, T_3$ for retrograde systems. For $\varepsilon^{23}, \varepsilon_{32} << 1$, $\alpha \simeq 1$, it follows that $T_2 \simeq T_3$. The mean motions are roughly equal in size but opposite in direction. If they were exactly equal, then $S = T_2/2 = T_3/2$ and consecutive conjunctions would occur exactly 180^o apart at opposite sides of the two binary orbits, (see Figures 4.4). Every second conjunction would occur in the same direction (with respect to the inertial frame). It is easily seen that

$$\frac{n_3^2}{n_2^2} = \frac{a_2^3}{a_3^3} \cdot \frac{G(m_1+m_2+m_3)}{G(m_1+m_2)} \doteq \alpha^3(1+\mu_3) = \alpha^3+\varepsilon_{32} \qquad (8)$$

For the systems described here, $\alpha^3 + \varepsilon_{32}$ is slightly less than one.

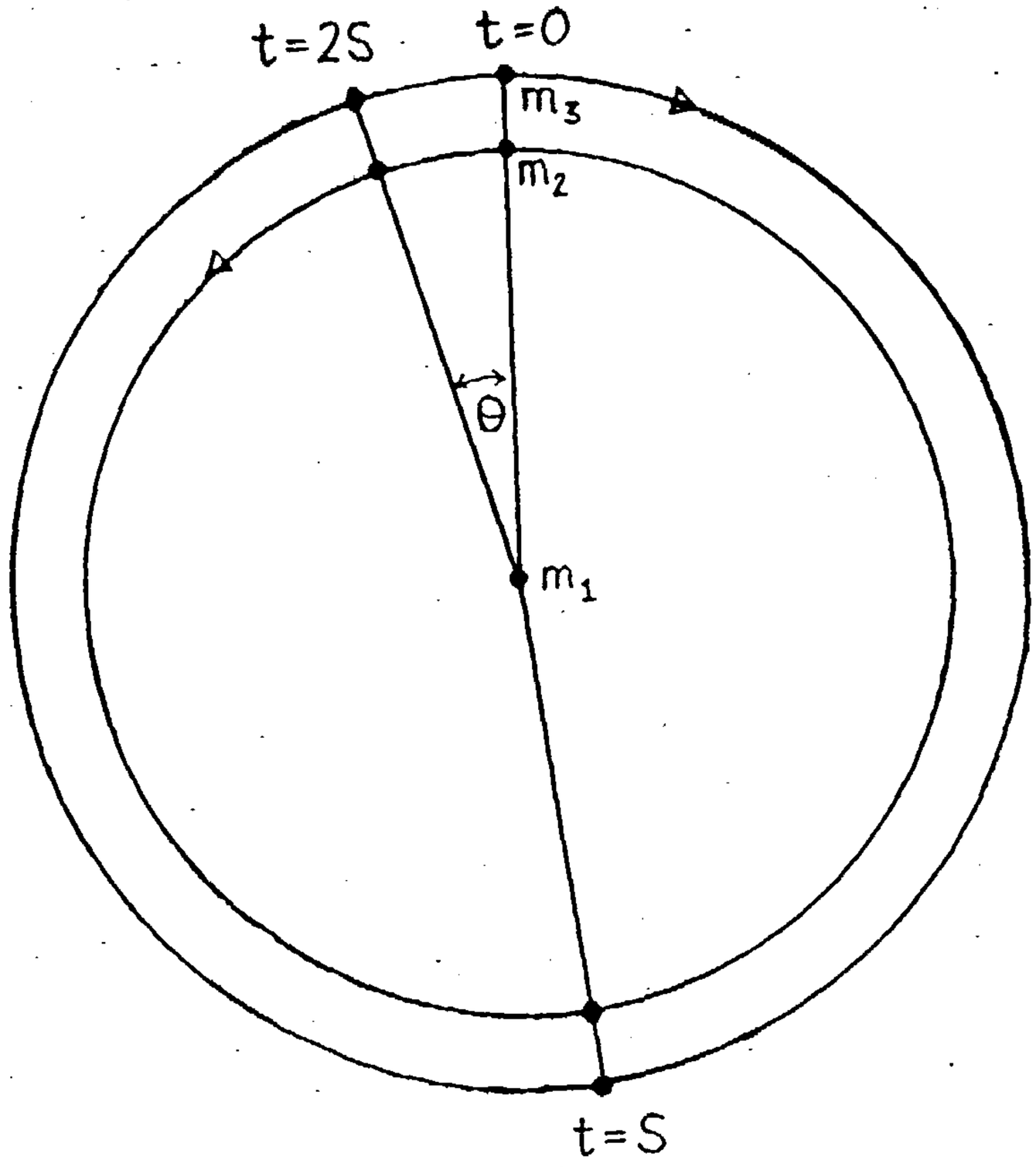


Figure 4.4: Retrograde system with similar binary periods. As $T_2 \to T_3$, $\theta \to 0^o$.

The two conjunction lines are seen to precess as in Figure 4.4 moving at a rate θ every two synodic periods

$$\theta = 2S.\ n_2 - 2\pi \qquad (9)$$

The precession period, assuming the elements remain fixed is equal to $2\pi.2S/\theta$ or in units of the synodic period is $4\pi/\theta$.

From Equations (7) and (9) it can be shown that

$$\frac{4\pi}{\theta} = \frac{2(1 + \frac{n_3}{n_2})}{(1 - \frac{n_3}{n_2})} \qquad (10)$$

where n_3/n_2 is given by Equation (8). Since $\varepsilon_{32} << 1$, $n_3/n_2 \simeq \alpha^{3/2}$. The precession period is quite sensitive to α , and may well be a factor in determining the period of the oscillation in $(a_2)_o$, $(a_2)_e$,

4.5

$(e_2)_o$, etc. It is seen that this period does decrease as α is decreased within the stable region. This cannot be the whole story as the observed changes are not as dramatic as Equation (10) would predict. It therefore seems likely that the period is caused by a more complicated dynamical mechanism which may come to light after deriving an analytical perturbation theory. This however is outside the scope of the present work.

To summarize, for given ε^{23} , ε_{32} , a system with initial $\alpha < \alpha_o$ seems to be stable for all time. $(a_2)_o$ and $(a_2)_e$ oscillate in anti-phase with common period, as do the odd and even conjunction values for a_3, e_2 and e_3 . As α is increased, the period of oscillation of the eight functions is seen to increase as do the average values until α reaches α_o. A secular term then appears which, over a period of time, causes the inner (m_1,m_2) binary orbit to cross the outer (M_2,m_3) binary orbit. As $\alpha(>\alpha_o)$ increases, the initial separation of the binary orbits decreases and the secular trend in a and e increases in strength, causing the time until crossover to decrease.

Recall in Section 4.4, there was an anomalous peak in the (α,N_s) curves for $\varepsilon_{32} = 10^{-2}$, $\varepsilon_{23} \leq 10^{-4}$. Figures 4.3(f),(g),(h) show a_2, a_3, e_2, e_3 for $\varepsilon^{23} = 10^{-6}$, $\varepsilon_{32} = 10^{-2}$. Figures 4.3(f) describes a system in the empirical stability region. Figure 4.3(g) describes an unstable system. Figure 4.3(h) describes a system which is stable for over 600 synodic periods, although its initial α-value exceeds that of the previous unstable system.

Comparing (f) and (h), we see that unlike the previous systems (a), (b), (c), (d), (e), there are distinct differences in the behaviour of the inner and outer elements. In particular, the outer orbit does

4.5

not show the same regularity in amplitude that the inner orbit does, although we must remember that the actual size of the oscillations are four orders of magnitude smaller. We noted previously that the period of oscillation of $(a_2)_o$, $(a_2)_e$, etc. increased as α increased. We note here that the period for system (h) is smaller than that of system (f).

For the previous systems as the initial α increased, the running value of a_2 increased and the running value of a_3 decreased (improving the chance of crossover). Comparing (f), (g), (h) we notice that as α is increased, the average value of a_3 falls as before, but so also does the average value of a_2.

A possible explanation for the enhancement of the lifetime of systems like system (h) may be that they suffer a very early close encounter. This close encounter rather than destroying the hierarchy has the surprising effect of changing the orbit so that the value of a_2 is lower than it otherwise would be thus lessening the chance of a crossover. This "constructive" close encounter only takes place over a certain range of α. It will be noted from Figures 4.1 that for α exceeding this range, the lifetimes are zero, indicating a strong "destructive" close encounter and subsequent crossover. Obviously this explanation is highly speculative and needs a much more detailed study to confirm or deny it.

4.6 Summary

A considerable amount of information has been presented in this chapter concerning retrograde three-body systems. A brief summary is useful at this stage.

In Section 4.5 a qualitative description was given of the general behaviour of the systems with the aid of graphs of osculating elements a_2, a_3, e_2, e_3. Various short and long period oscillations in these elements were pointed out which indicate that an unstable system changes slowly, with the binary orbits approaching each other and finally crossing, changing the hierarchy. The lifetime is a function of the relative sizes of the epsilons and the initial separation of the orbits as given by α. Some of the explanations given for the observed behaviour cannot be justified without more study, possibly studying the elements of other systems by numerical integration, as before, or employing analytical perturbation techniques to highlight interesting periodicities.

Section 4.4 gave the results from several hundred numerical integrations of different systems by graphs of lifetime against α for given pairs of epsilons. Although for many of the epsilon pairs there was no analytical guarantee of stability for any α, there always seems to exist a value α_o below which no unstable systems could be found. For $\alpha > \alpha_o$, the systems were seen to be unstable and the stability lifetimes were seen to decrease monotonically (with the exceptions of $\varepsilon_{32} = 10^{-2}$, $\varepsilon^{23} \lesssim 10^{-4}$ where the singular peaks in the lifetimes occurred). The range of α over which the lifetimes fell from over 600 synodic periods (integration limit) to one synodic period varied with the epsilons. For $\varepsilon^{23} < \varepsilon_{32}$, smooth analytical curves were fitted to the data over this range. For $\varepsilon^{23} \gtrsim \varepsilon_{32}$, the range was too small for a curve to be usefully fitted. In these cases only α_o was determined.

It is worth summarising the curve fit procedure for $\varepsilon^{23} < \varepsilon_{32}$.

We are fitting the curve

$$N_s = f(\alpha) = \exp\left[\beta\left(\frac{1-\alpha}{\alpha-\alpha_o}\right)^{\gamma}\right] - 1 \quad , \quad \alpha_o < \alpha \leq 1.$$

characterised by the parameters α_o, β, γ , for given ϵ^{23} , ϵ_{32} .

We wish to weight this curve fit towards the low α-high N_s data.

Therefore for each step of data points (i.e. data with common lifetime) discard all the points except one in the middle of the step. Weight the remaining data points according to their N_s values. Calculate quantities u,v as given in Equations (4) and (5) for a chosen value of α_o. Determine by least squares applied to the (u,v) data, the corresponding values of β and γ . We have found the best curve that fits the unstable system data for a chosen α_o. It must also fit the "stable" data in the sense that it should indicate any system that will survive for over 600 synodic periods. If it fails to do this for any of the observed systems, the curve should be discarded and this value of α_o disallowed. We pick the curve that gives an allowed value of α_o (with corresponding β and γ) that maximises the correlation coefficient r as given in Equation (6).

We saw that the fitted curves, although showing a systematic error, in most cases gave a fairly accurate prediction of the actual stability lifetime for any given system.

We cannot say for certain that the value of α_o that we have found is the value of α such that all systems, with initial $\alpha < \alpha_o$, are stable for all time. We can say this however, insofar as $N_s = f(\alpha)$ is an accurate function for fitting to the (α, N_s)data, then there is a region of empirical stability for $\alpha_c < \alpha < \alpha_o$, where α_o is substantially greater than the value α_c, calculated by the analytical

4.6

stability theory.

There is no reason why this curve-fit procedure should not be applied to the direct systems. The use of the correlation coefficient excludes the possibility of subjective bias in determining α_o, as was the case in Walker and Roy, Paper III. Our success in being able to predict the stability lifetimes for retrograde systems encourages us to do the same for direct systems. As we shall see in Chapter 5, this will turn out to be a much more difficult task.

CHAPTER 5

PREDICTIONS OF STABILITY LIFETIMES FOR DIRECT SYSTEMS

5.1 Introduction

This chapter is concerned with applying statistical techniques to data from coplanar direct three-body systems in order to predict the stability lifetimes of such systems. The techniques described here are based on those described in the previous chapter as well as the work of Walker and Roy. Inevitably there will be comparisons made between their methods and the augmented methods given here.

In the same manner as for the retrograde systems, using the same integration routine described in Section 4.3, approximately 450 direct three-body systems were studied. All were coplanar, initially circular, starting from the m_1-m_2-m_3 conjunction configuration as before. Recall that these systems are uniquely defined by ε^{23}, ε_{32}, α . All the systems took values

$$\varepsilon^{23} = 10^{-i}$$
$$\varepsilon_{32} = 10^{-j} \qquad i,j = 2,3,4,5,6.$$

Of the 450 systems investigated, 150 were taken from the original data set of Walker and Roy, Paper III. Some of their systems were reintegrated as the diagnostics within the integration program for detecting close encounters had been improved. Checks now take place after every step instead of every conjunction. This is particularly useful for detecting close encounters for systems with stability lifetimes of 5 synodic periods or less. Whereas Walker had integrated systems for a total of 500 synodic periods, this limit has been raised to 4000 synodic periods. Accumulated integration error prevent us integrating further, (Section 4.3). As a result many systems from the original data set which were stable over 500 synodic periods, exhibit instabilities after longer times.

5.1

As was said above, this chapter is primarily concerned with data-processing. The dynamical implications of this processing are largely left to Chapter 6 where the results from the fictitious systems of Chapters 4 and 5 are compared with real systems. The behaviour of direct three-body systems is discussed very fully in Papers I, II and III by Walker and Roy and also in Walker (1980). Some relevant results are described in Section 5.2 along with graphs of stability lifetime against α for different values of ε^{23} , ε_{32} taken from the 450 integrated systems.

Section 5.3 is concerned with fitting curves to the (α, N_s) data in a similar way to the retrograde systems. We shall see that curve-fitting is made more difficult, because commensurabilities cause a considerable spread in the data. These effects are discussed in Section 5.4.

In Section 5.5, a method is found for normalising the (α, N_s) data for each pair of epsilons. This allows all the available data to be used together in order to obtain a quantitative evaluation of the spread of lifetimes about a predicted value. This allows us to test our procedures against a new set of 100 three-body systems to see if their behaviour matches our predictions.

5.2 Numerical Experiments for Direct Three-Body Systems

The most striking aspect concerning direct three-body systems is the variety of behaviour that they exhibit. The retrograde systems were fairly consistent in the way they evolved either in a stable or unstable manner (Section 4.5). In the case of direct systems, some show steadily rising eccentricities which lead inevitably to a crossover of orbits (as for retrograde systems). On the other hand, many more show

5.2

osculating eccentricities which vary periodically without any sign of a secular increase until a "close encounter" takes place. The eccentricities rise sharply to resume their oscillations, this time about a new mean value, with a different amplitude and in some cases with a different period. The new amplitude need not be greater than before, but may in fact be smaller. In extreme cases, the eccentricities may be nearly constant. This may indicate that a system has been perturbed into a stable commensurable configuration.

The method for diagnosing close encounters when the osculating elements of the inner and outer binary orbits change dramatically was described fully in Section 4.3. On occasions, a system will show a sudden rise in eccentricities which looks irreversible but after a longer time, these eccentricities will fall again. We are then faced with a dilemma. Do we treat this as a long period trend of the original system or as two close encounters resulting in two drastic changes in the system? The stability lifetime will depend on the interpretation that we attach to the data. This is one of the problems encountered in interpreting the behaviour of direct systems: we are unable to avoid a subjective analysis. Faced with the situation above, how do we judge the irreversibility of changes in eccentricity? If we choose to ignore close encounters altogether, as we did with the retrograde systems, and use only the objective tests of escape and crossover, then we miss detecting many systems that evolve into a more stable configuration which they can maintain for all time.

In the event, we choose to retain the close encounter diagnostic and accept the subjectivity. Systems which exhibit very long period changes are considered to be stable, unless proved otherwise. In any case, the

5.2

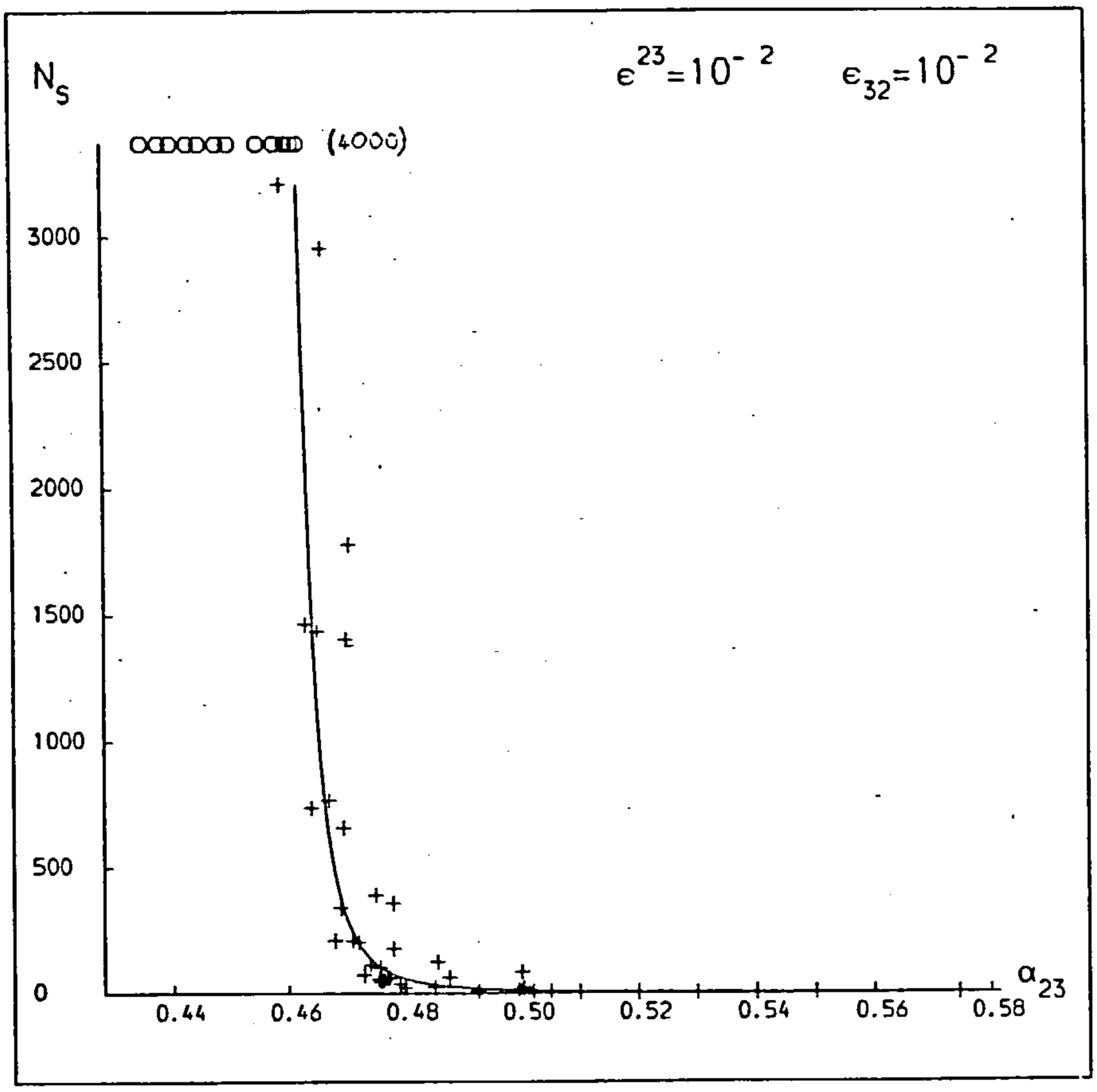
number of ambiguous systems is relatively small. For most systems there is little doubt that the changes are irreversible. For example, a set of close encounters is often succeeded by a crossover.

There is a further ambiguity in classifying the type of instability even if we know it has definitely taken place. Did the crossover arise as a result of the close encounters or were the close encounters a symptom of the approaching crossover. Generally the former is assumed here and the stability lifetime is taken to be the time to close encounter. In most cases of this type, the difference in time is very small so the actual value of the lifetime will not be badly affected. Walker and Roy chose to include the classification of the method of instability in Paper III. However, for reasons of ambiguity and because it is not relevant to the subsequent work, we shall ignore such a classification.

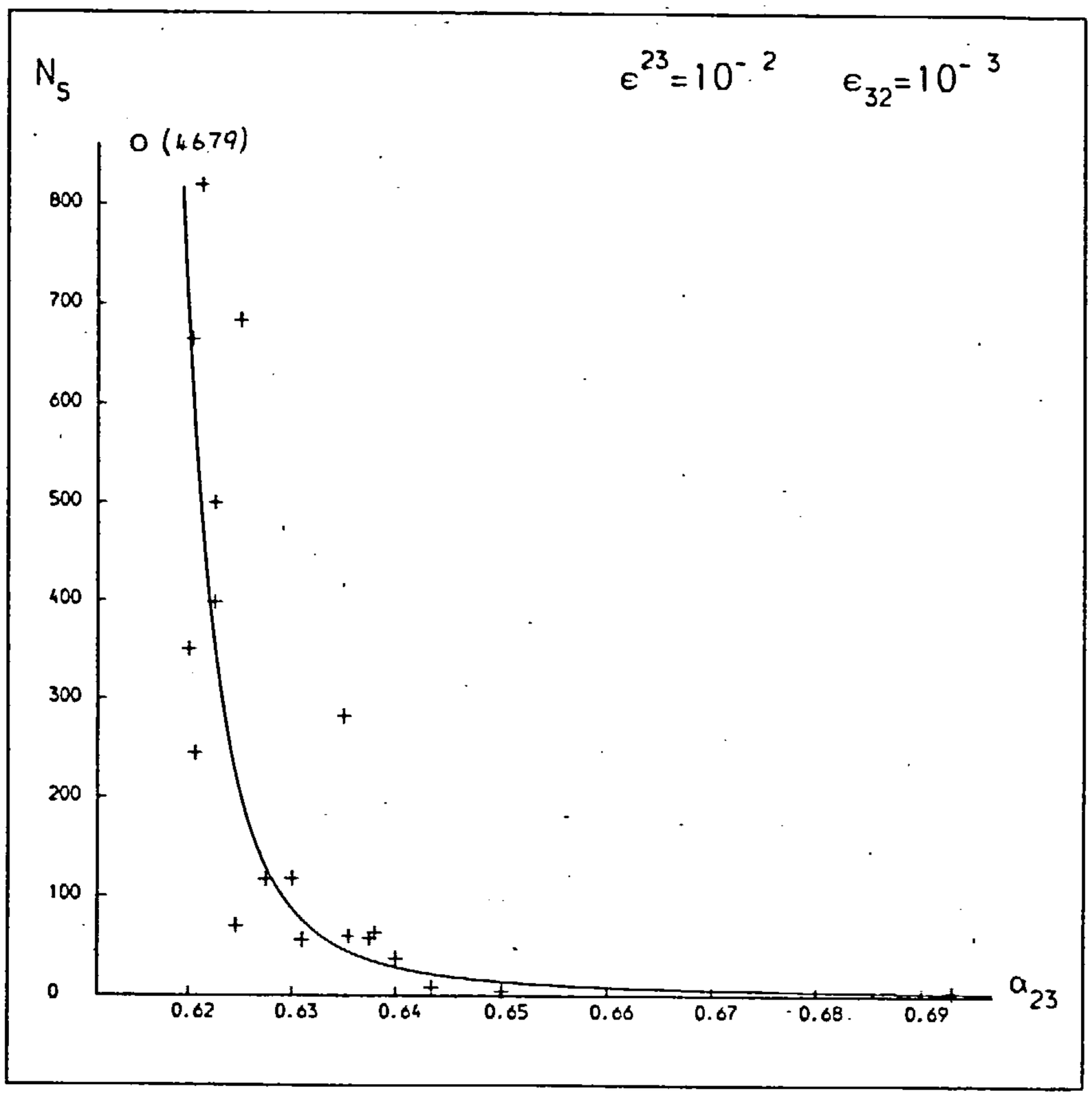
It can be seen as a general trend that the amplitudes of variation of the orbital elements are proportional to ε^{23} , ε_{32} in the same manner as for the retrograde systems (Section 4.5). Walker (1980) has done extensive investigations into this and it is not proposed to repeat them here. Suffice it to say, that this proportionality is a reflection of the definitions of ε^{23} and ε_{32} as a measure of the relative perturbation on the outer and inner binary orbits, respectively.

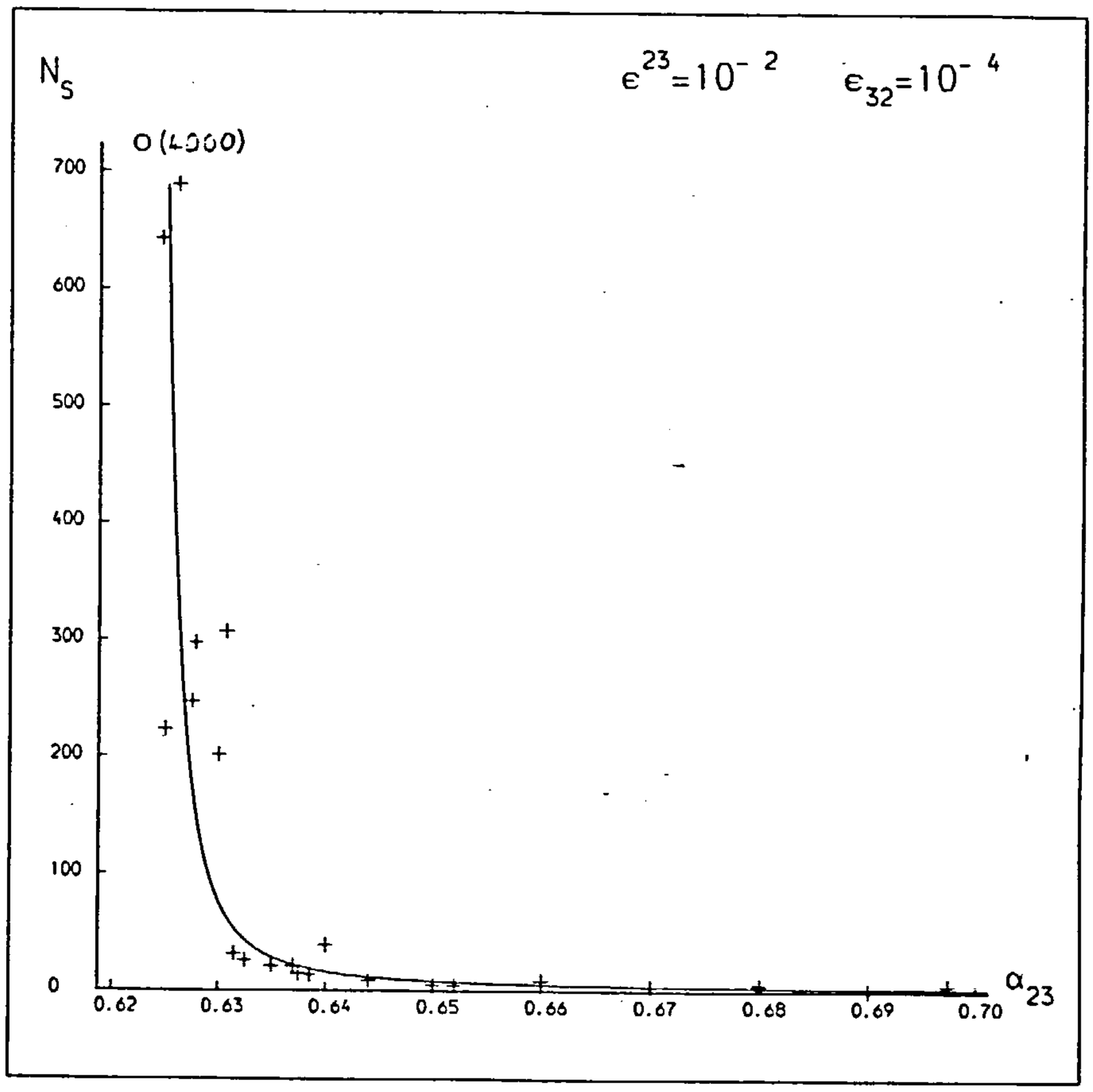
It can also be confirmed that on the few occasions when one of the bodies escaped the system, it was always the least massive. Marchal (1985) has shown that this need not always be the case, but that it is the most likely.

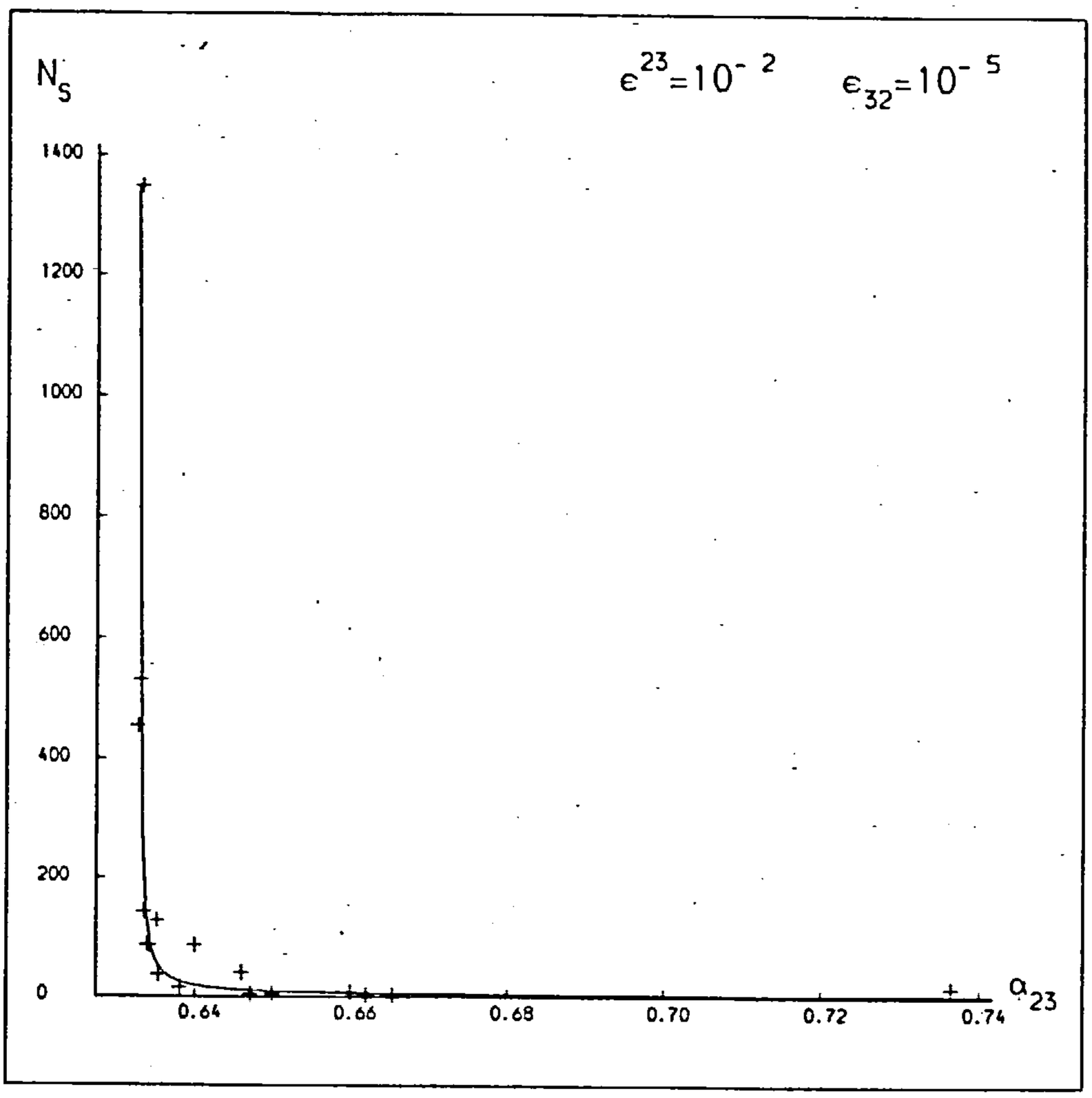
The above comprises only a summary of some aspects concerning the behaviour of direct three-body systems, a more detailed analysis being given by Walker (1980). It is however sufficient as a background

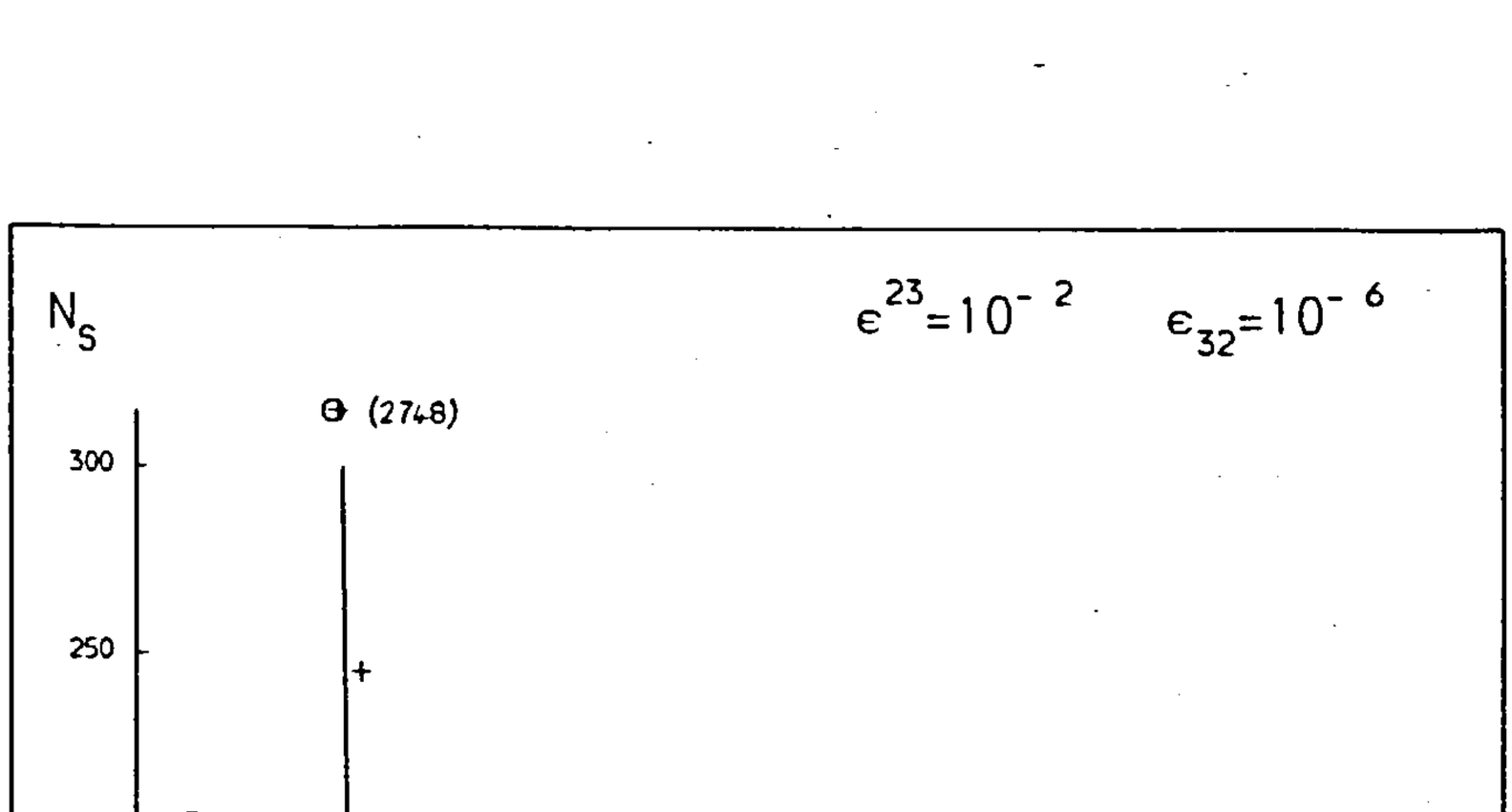


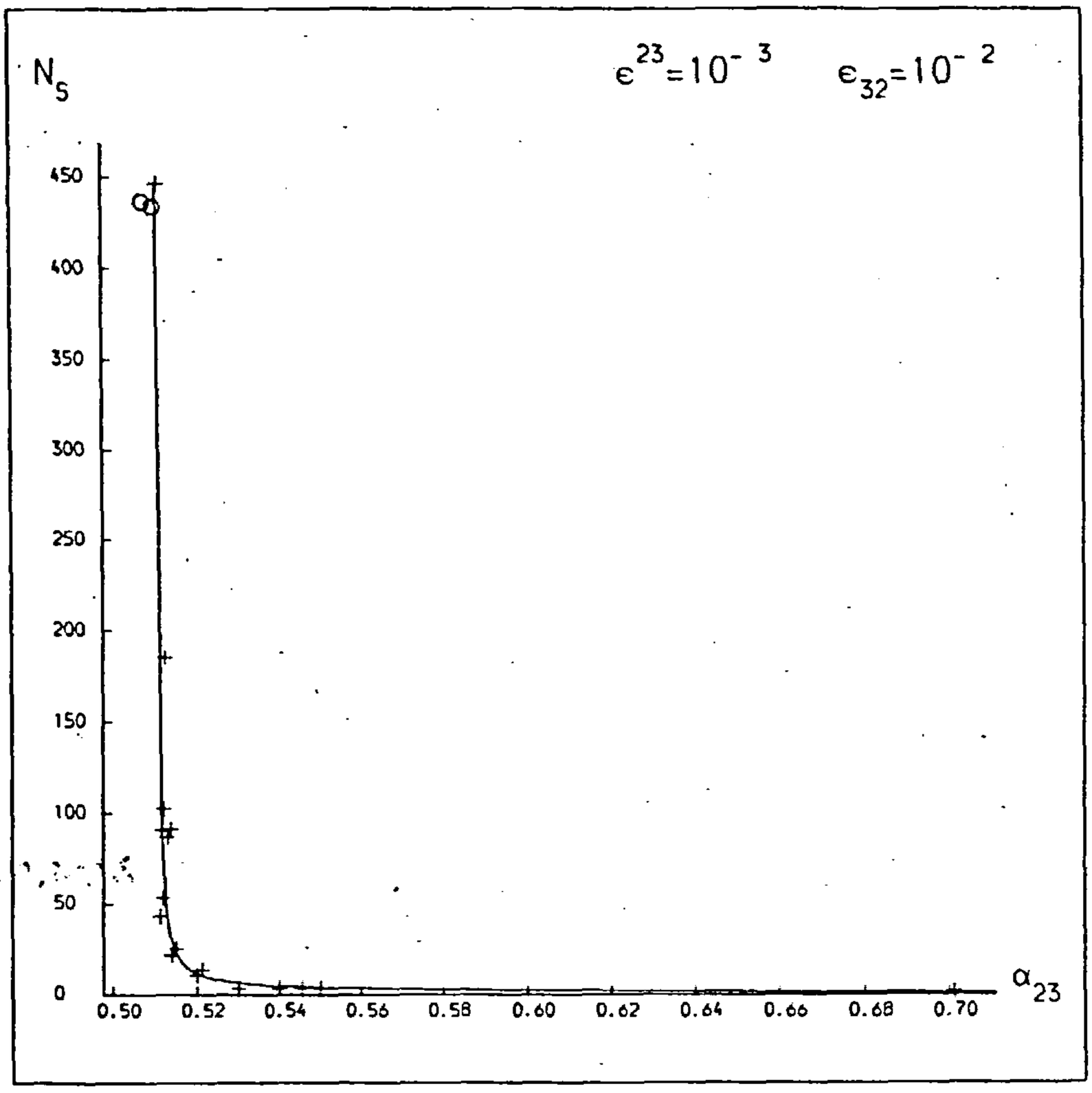
Figures 5.1: Stability lifetime against α for given ϵ^{23} , ϵ_{32} for direct systems. The crosses indicate the stability lifetime for each unstable system. The circles indicate systems that showed no signs of instability during the given lifetime. A best fit curve is drawn through the unstable points. Where some systems have lifetimes exceeding the graphs limits, the value of the lifetime is given in brackets.

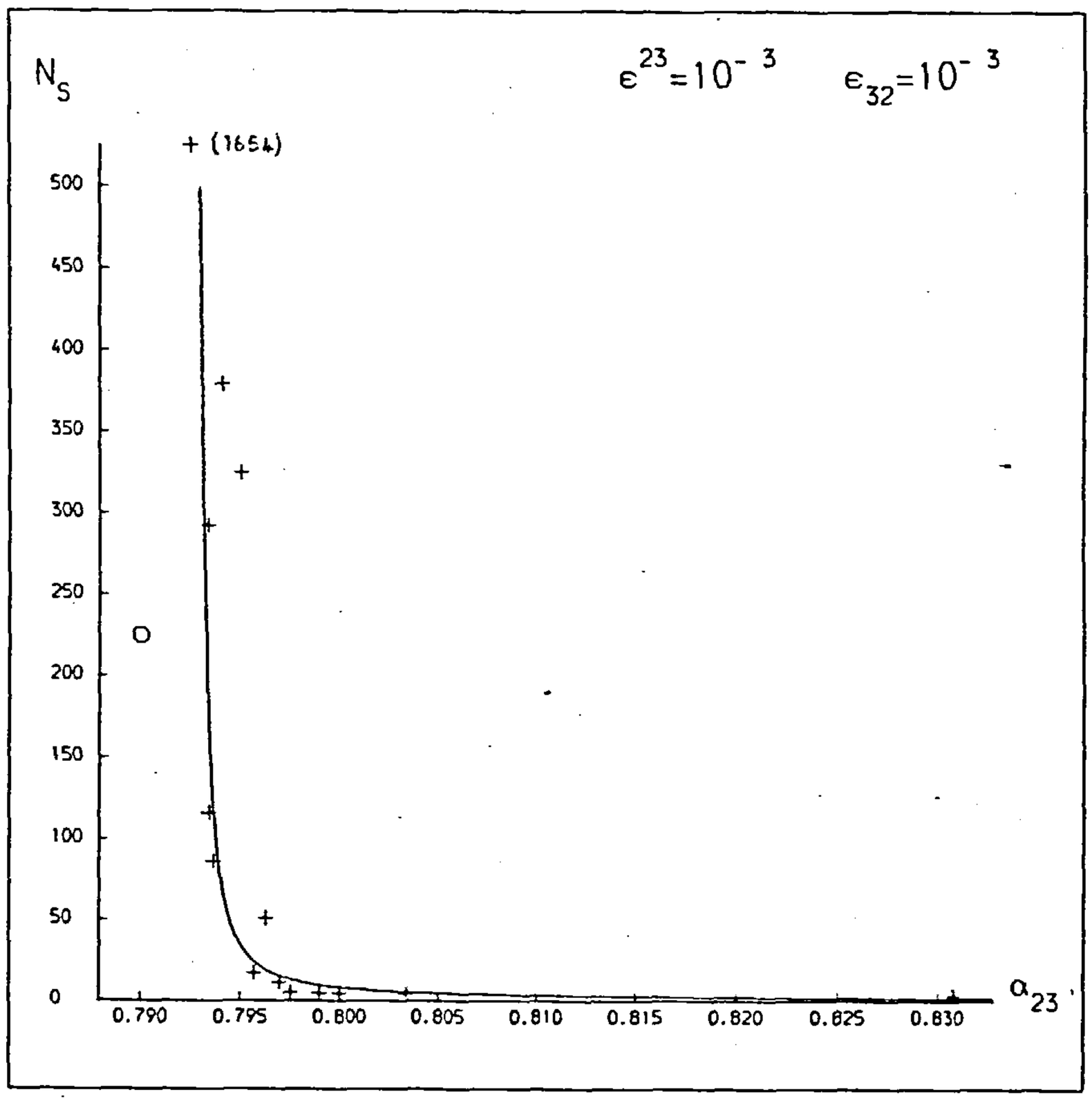

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-3}$
O (4679)
800
700
600
500
400
300
200
100
0
0.62
0.63
0.64
0.65
0.66
0.67
0.68
0.69
a_{23}

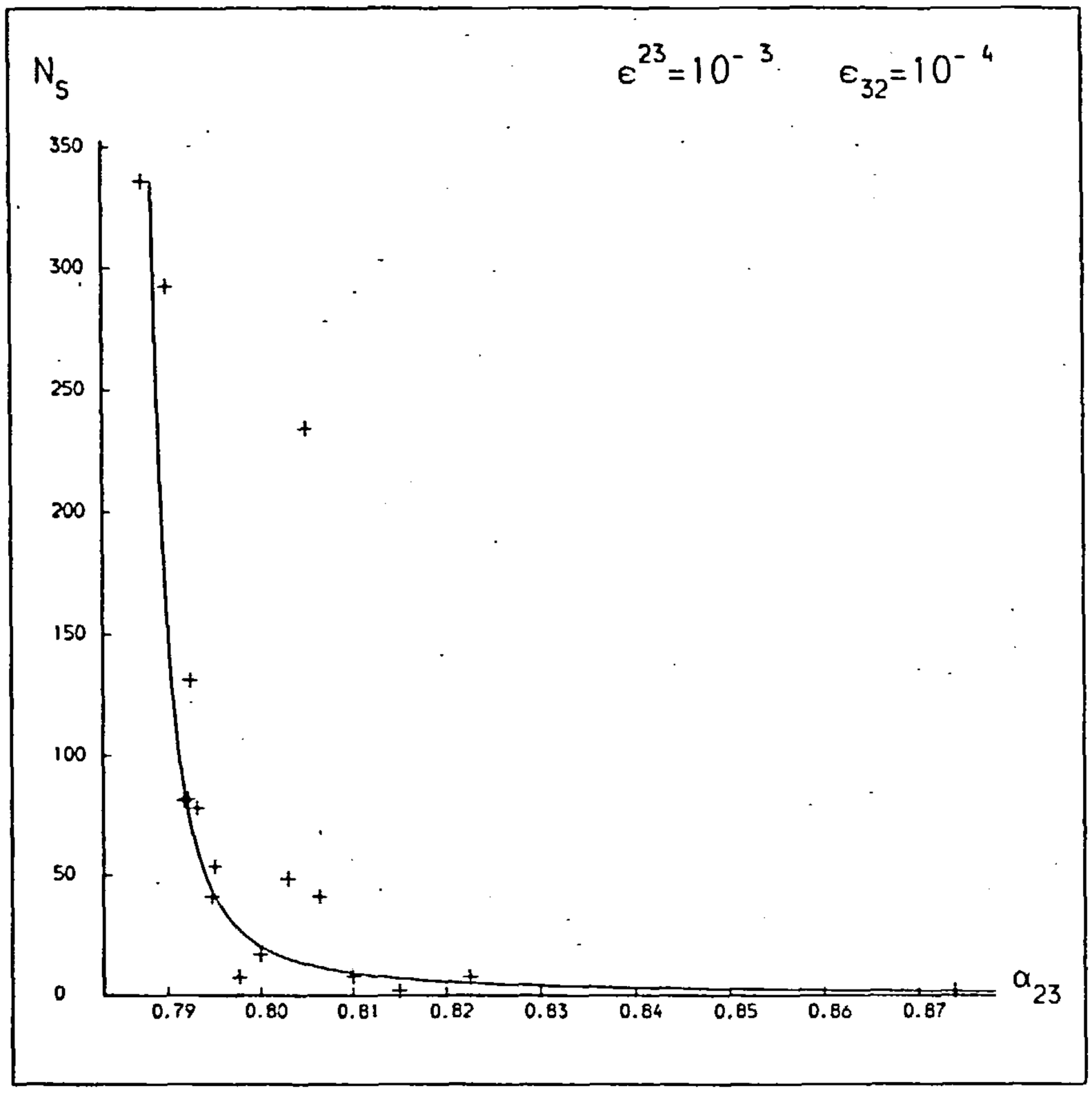

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-4}$
O (4000)
700
600
500
400
300
200
100
0
0.62
0.63
0.64
0.65
0.66
0.67
0.68
0.69
0.70
a_{23}

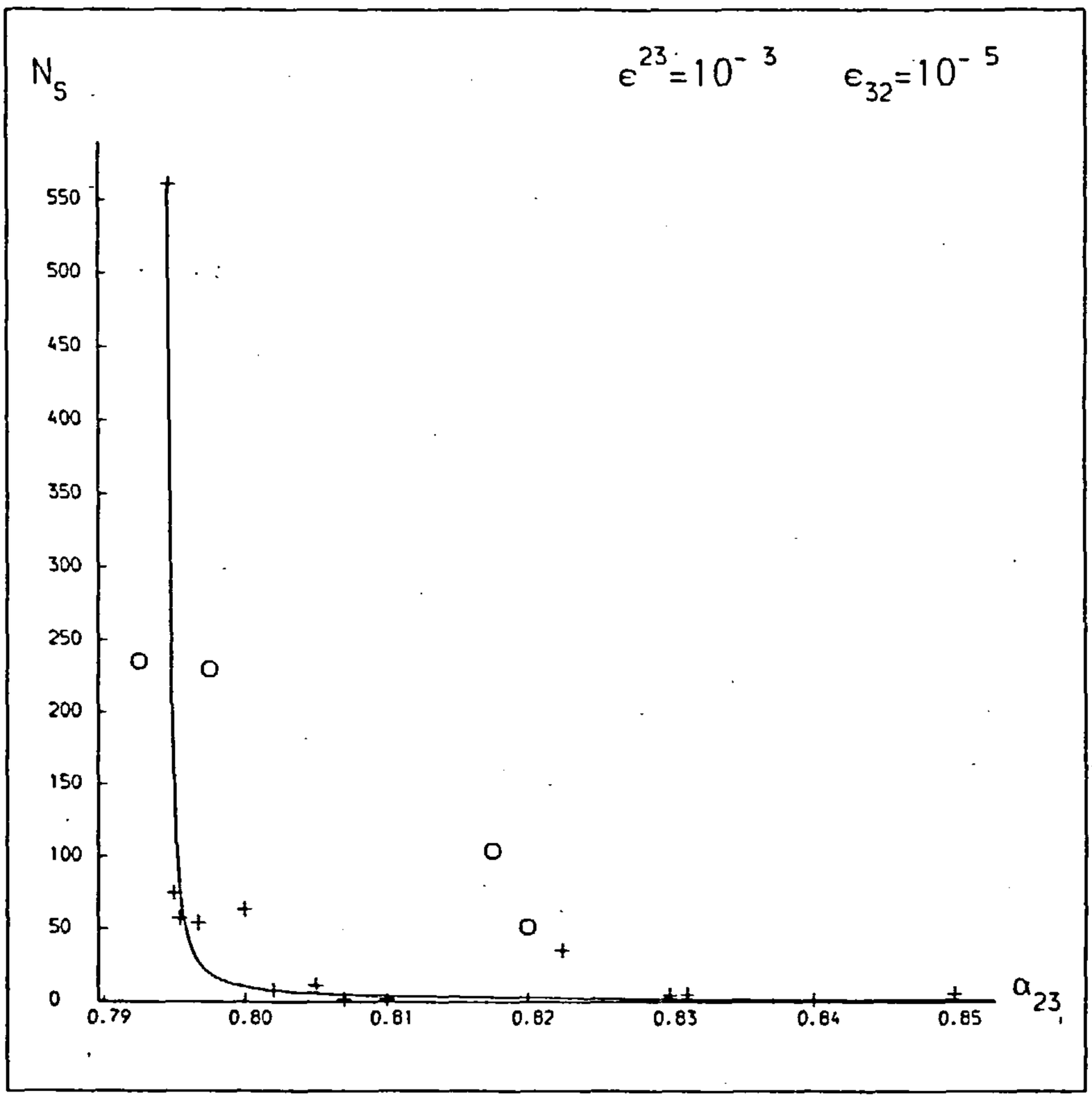

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-5}$
1400
1200
1000
800
600
400
200
0
0.64
0.66
0.68
0.70
0.72
0.74
α_{23}

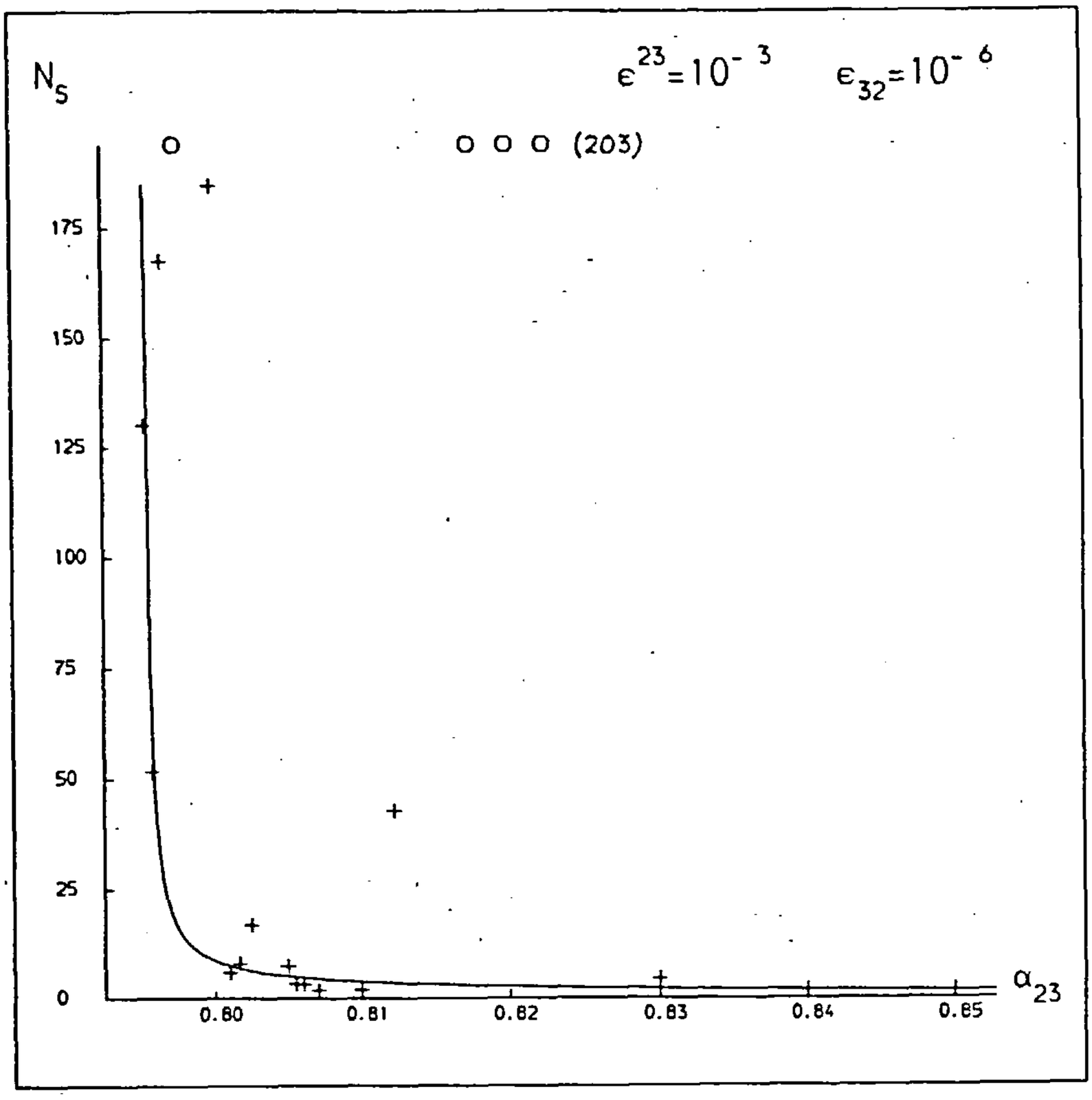

N_S
$\epsilon^{23}=10^{-2}$
$\epsilon_{32}=10^{-6}$
Θ (2748)
300
250
200
150
100
50
0
0.62
0.63
0.64
0.65
0.66
0.67
0.68
α_{23}

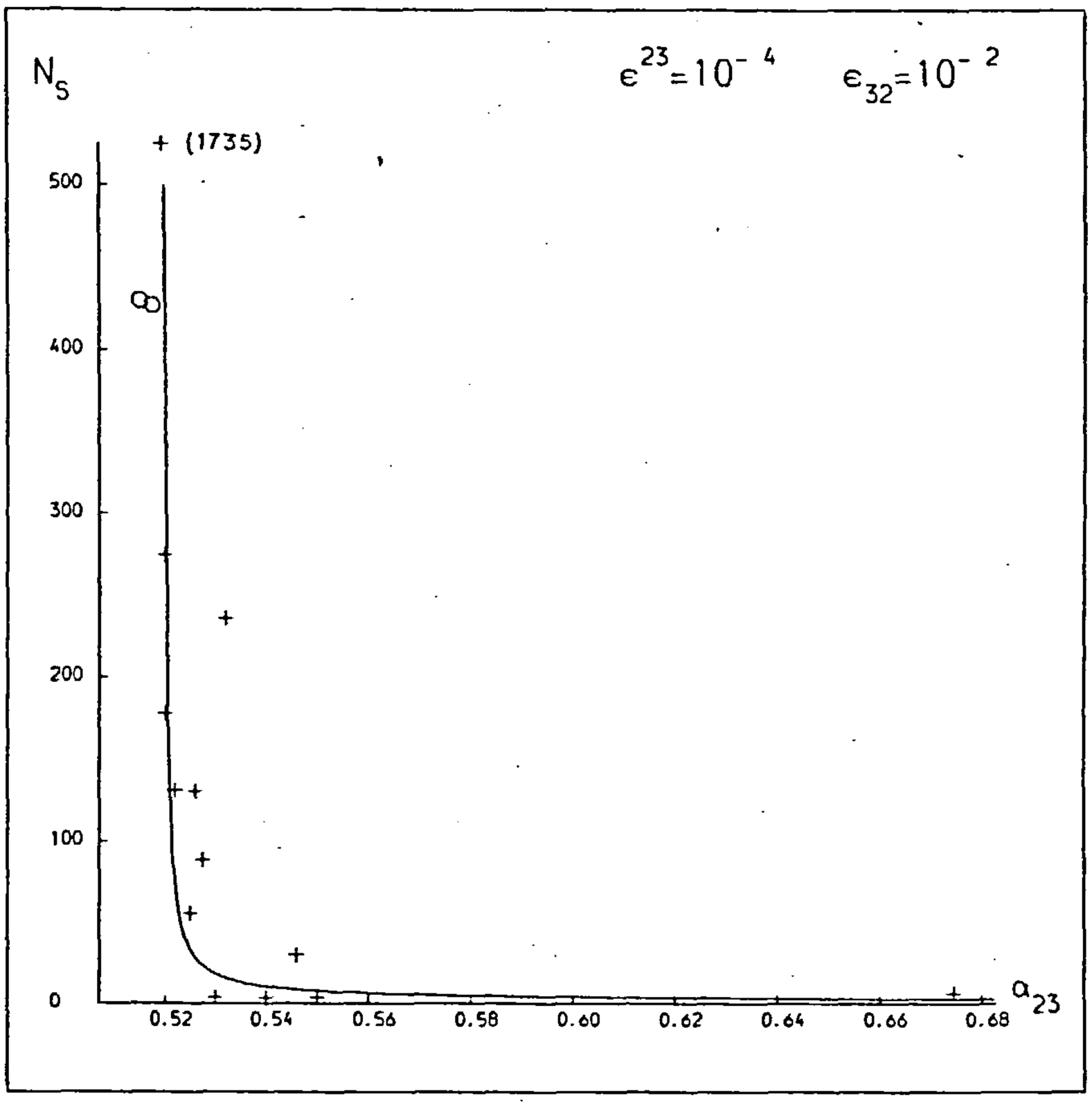

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-2}$
450
400
350
300
250
200
150
100
50
0
0.50
0.52
0.54
0.56
0.58
0.60
0.62
0.64
0.66
0.68
0.70
α_{23}

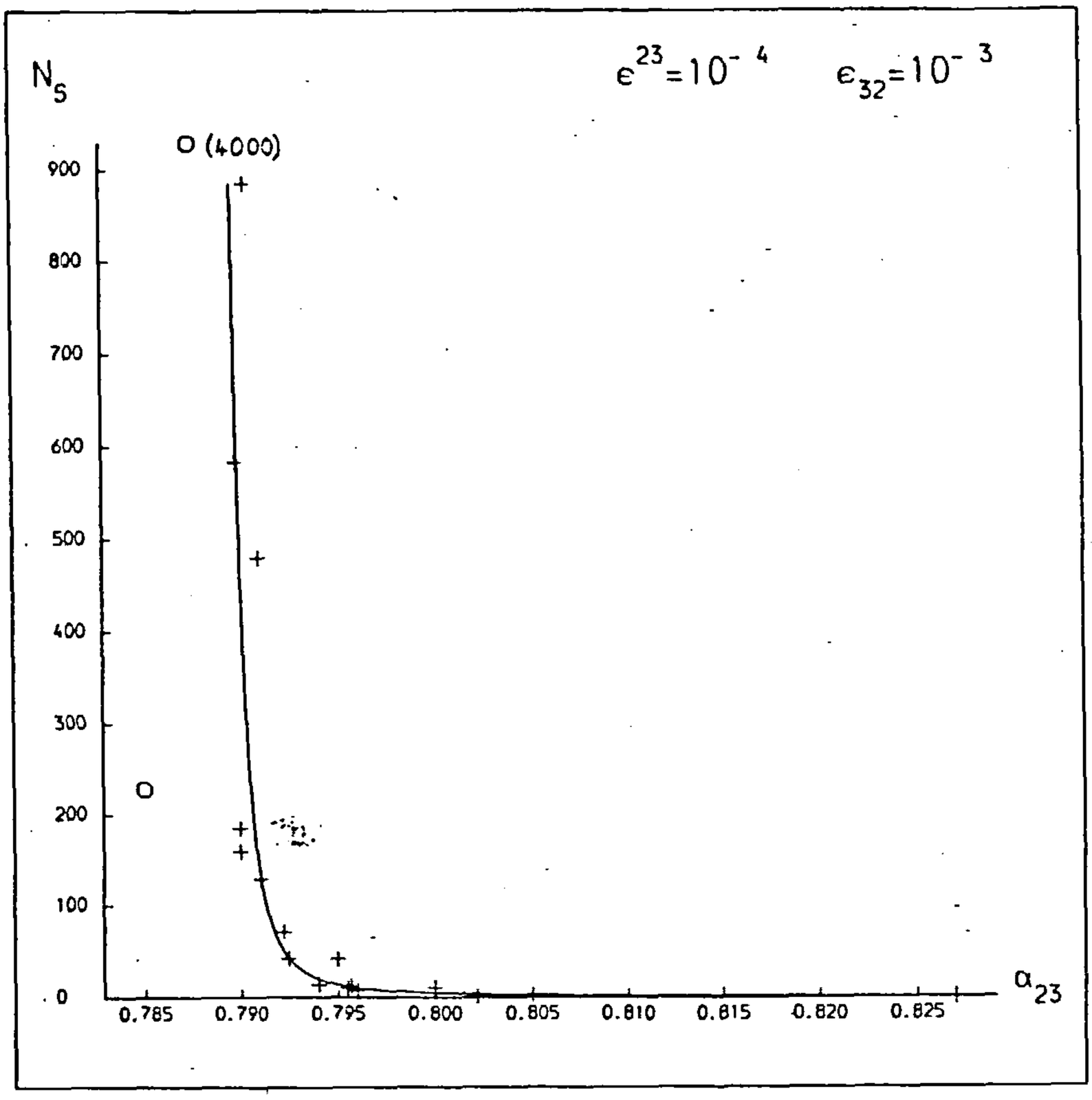

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-3}$
+ (1654)
500
450
400
350
300
250
200
150
100
50
0
0.790
0.795
0.800
0.805
0.810
0.815
0.820
0.825
0.830
α_{23}

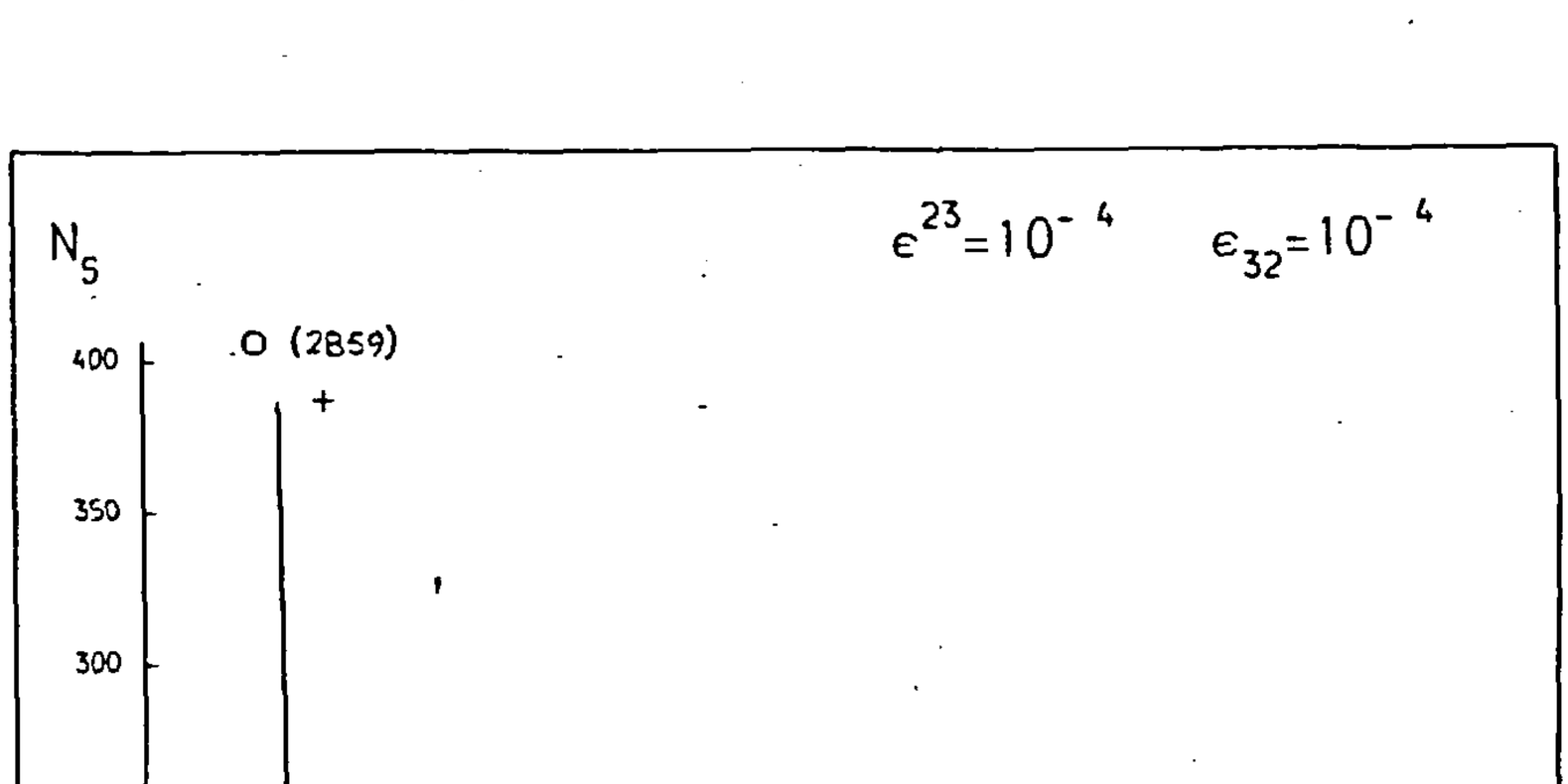

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-4}$
350
300
250
200
150
100
50
0
0.79
0.80
0.81
0.82
0.83
0.84
0.85
0.86
0.87
α_{23}

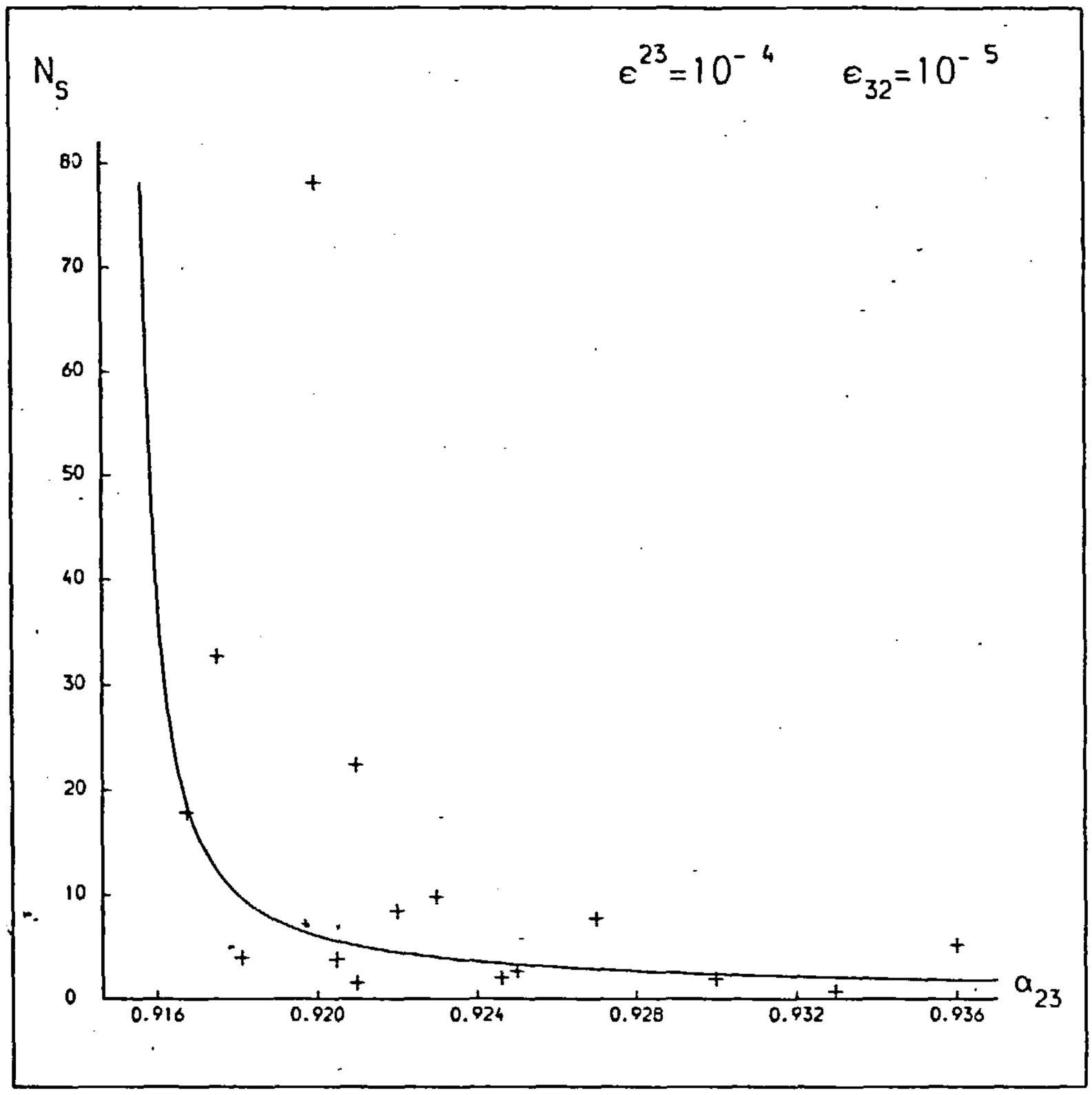

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-5}$
550
500
450
400
350
300
250
200
150
100
50
0
0.79
0.80
0.81
0.82
0.83
0.84
0.85
α_{23}

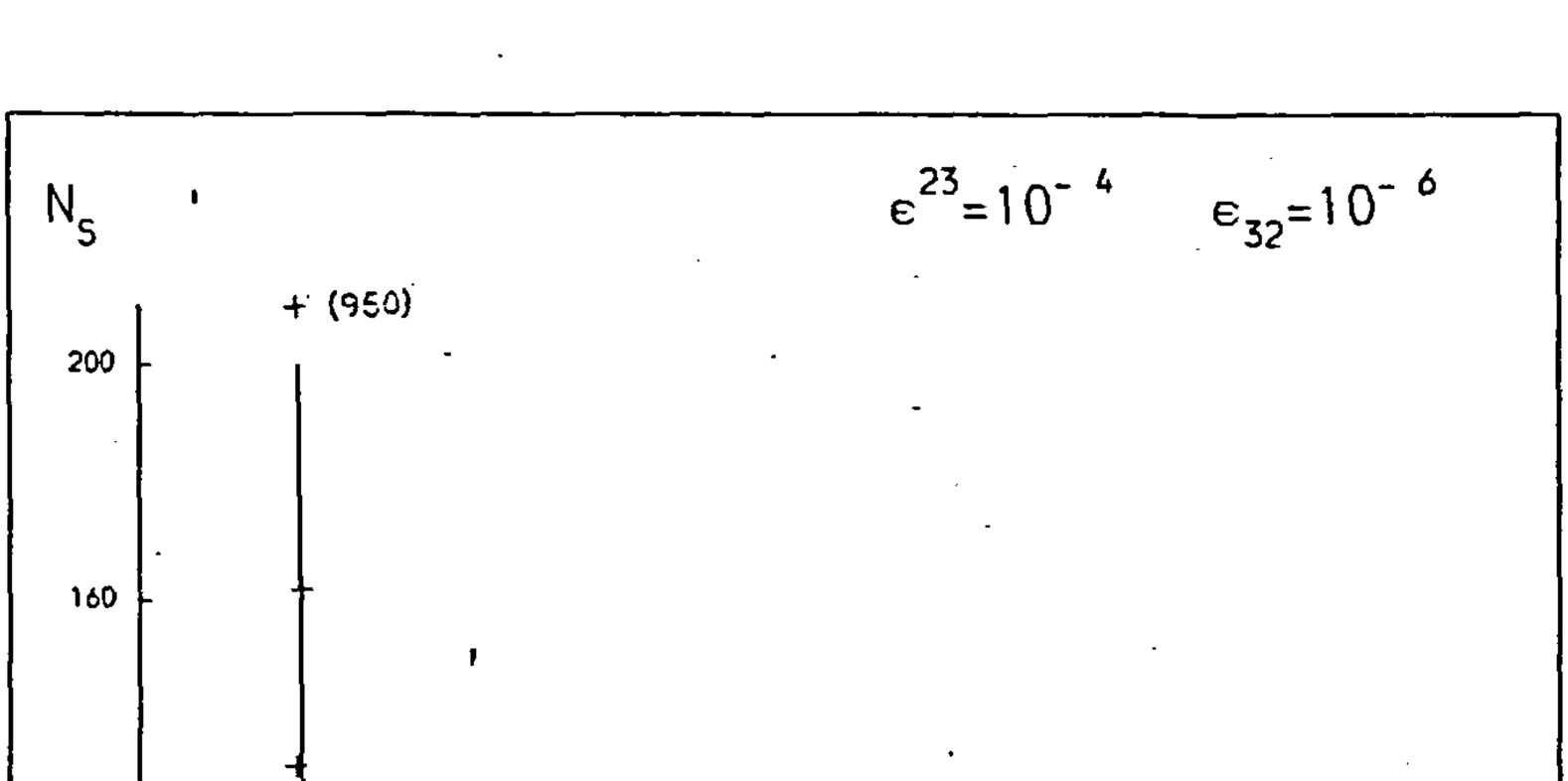

N_S
$\epsilon^{23}=10^{-3}$
$\epsilon_{32}=10^{-6}$
o o o (203)
175
150
125
100
75
50
25
0
0.80
0.81
0.82
0.83
0.84
0.85
α_{23}

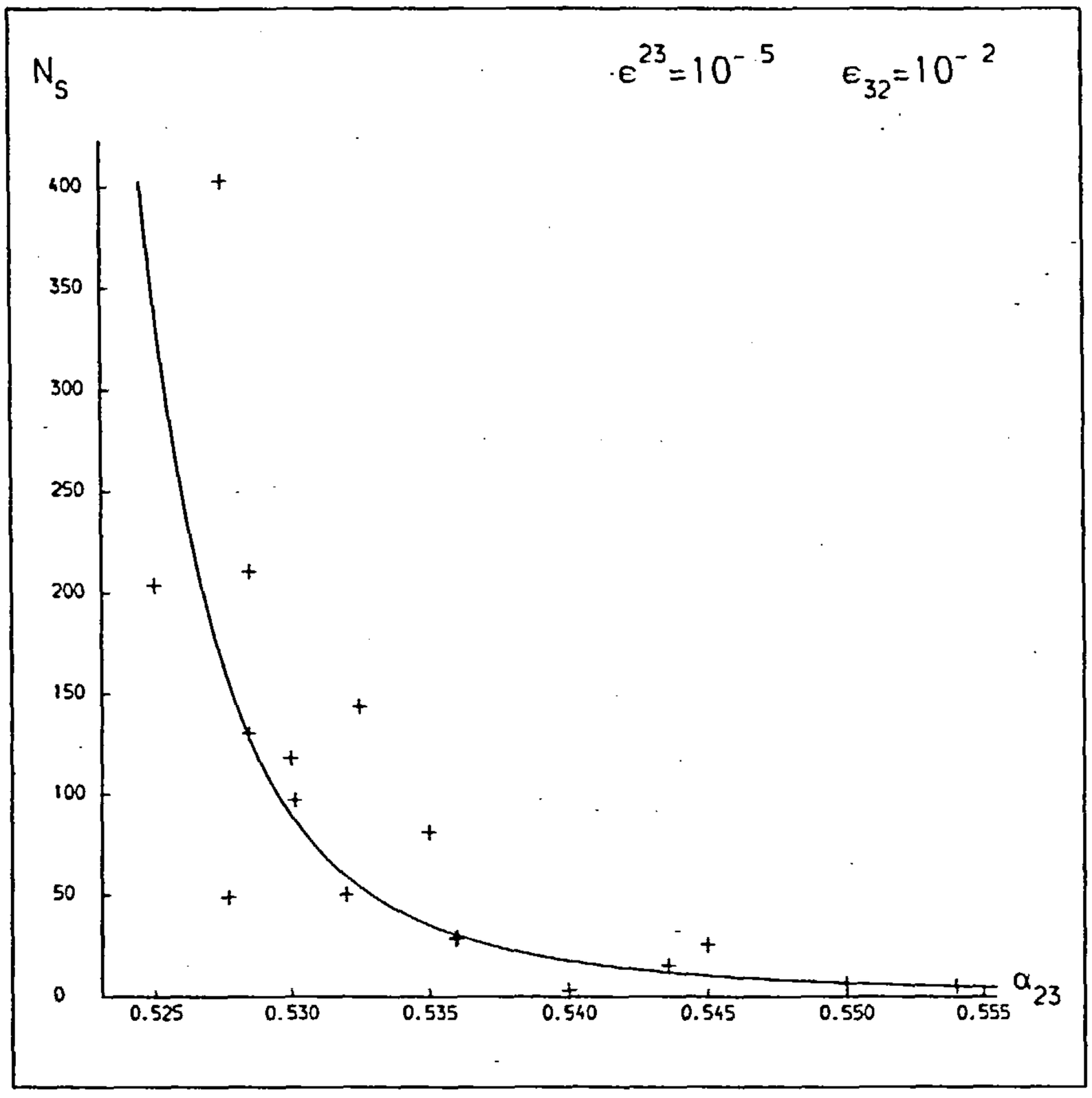

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-2}$
+ (1735)
500
400
300
200
100
0
0.52
0.54
0.56
0.58
0.60
0.62
0.64
0.66
0.68
α_{23}

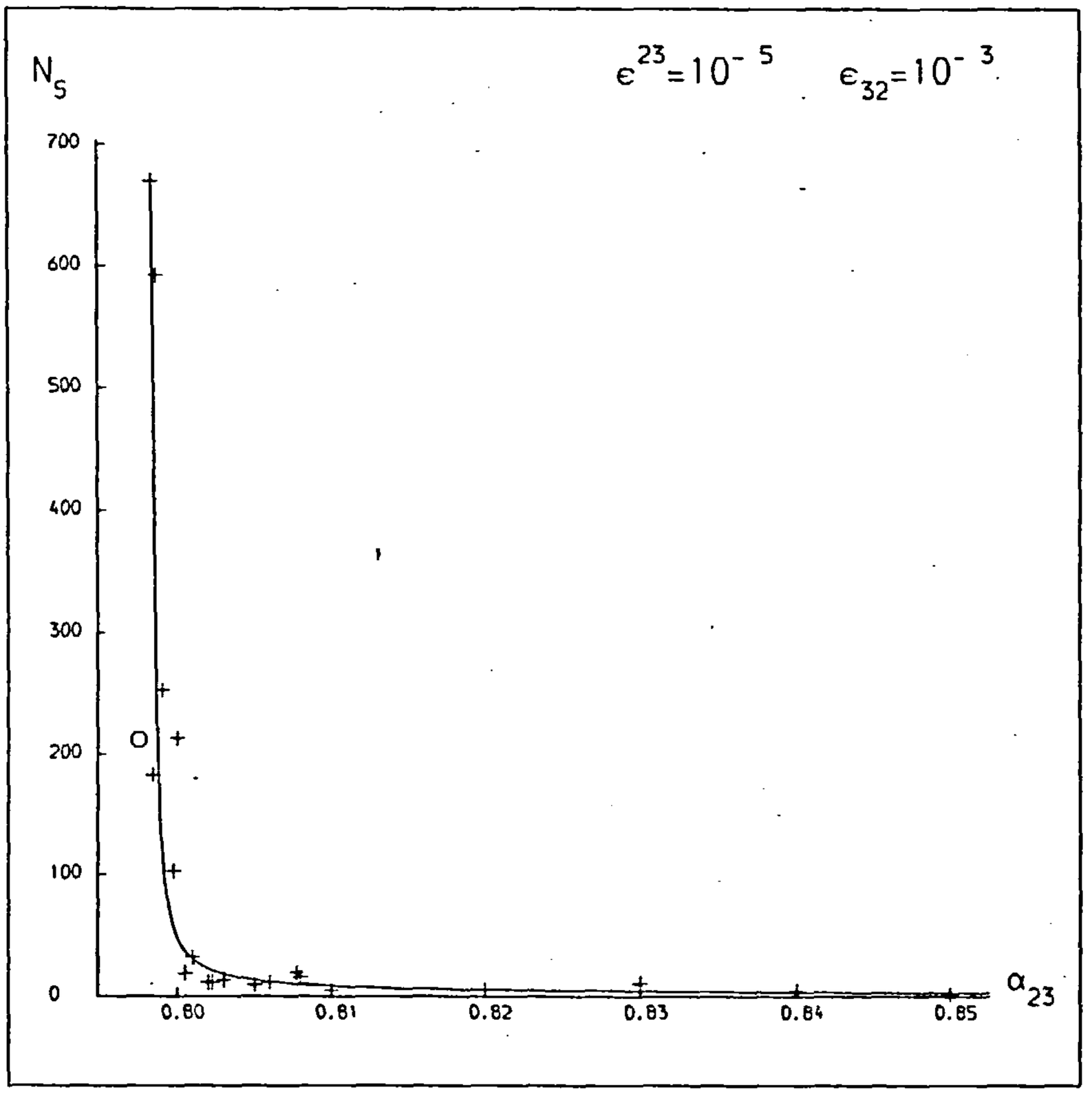

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-3}$
O (4000)
900
800
700
600
500
400
300
200
100
0
O
0.785
0.790
0.795
0.800
0.805
0.810
0.815
0.820
0.825
α_{23}

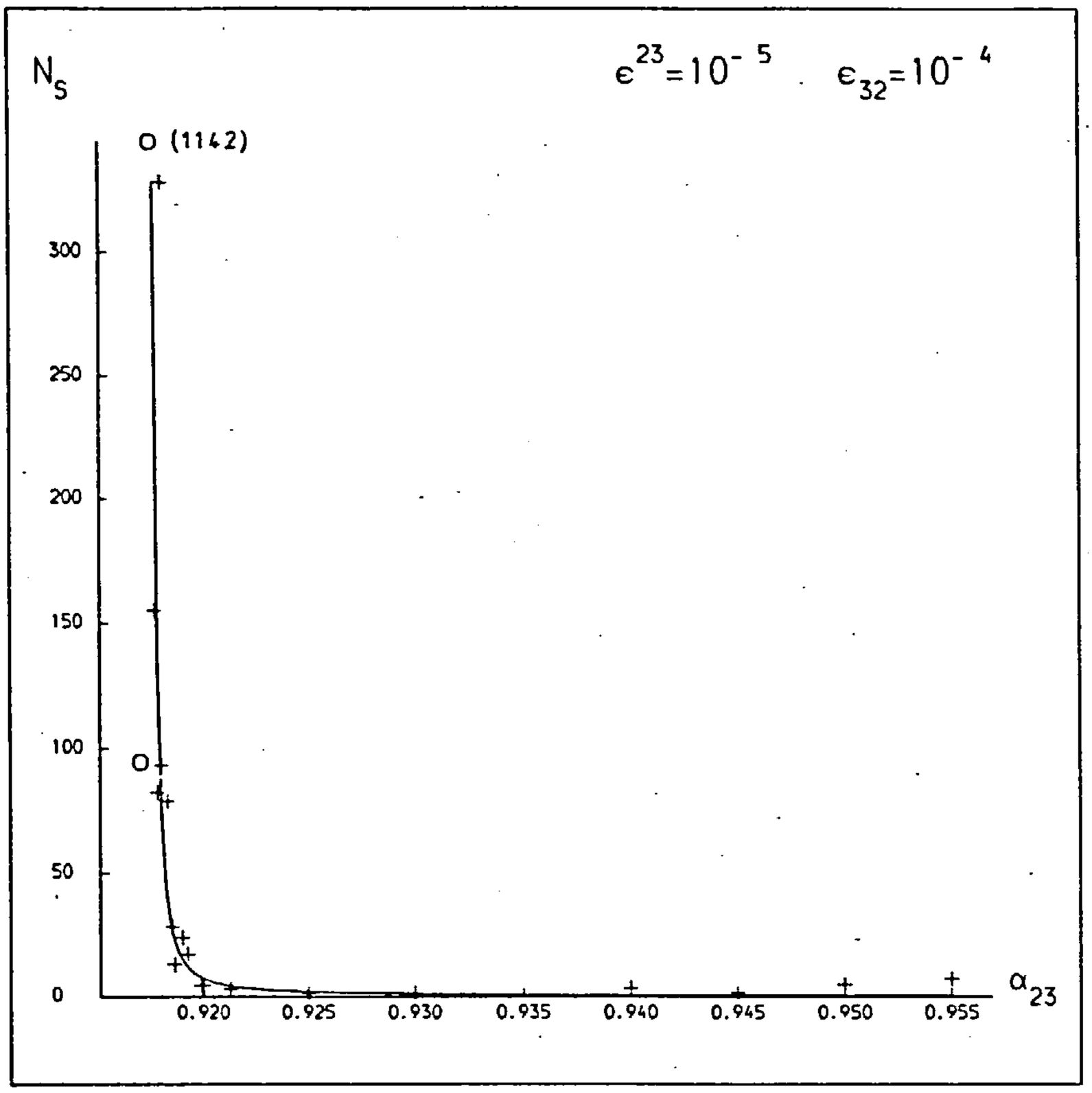

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-4}$
O (2859)
400
350
300
250
200
150
100
50
0
O
0.90
0.91
0.92
0.93
0.94
0.95
α_{23}

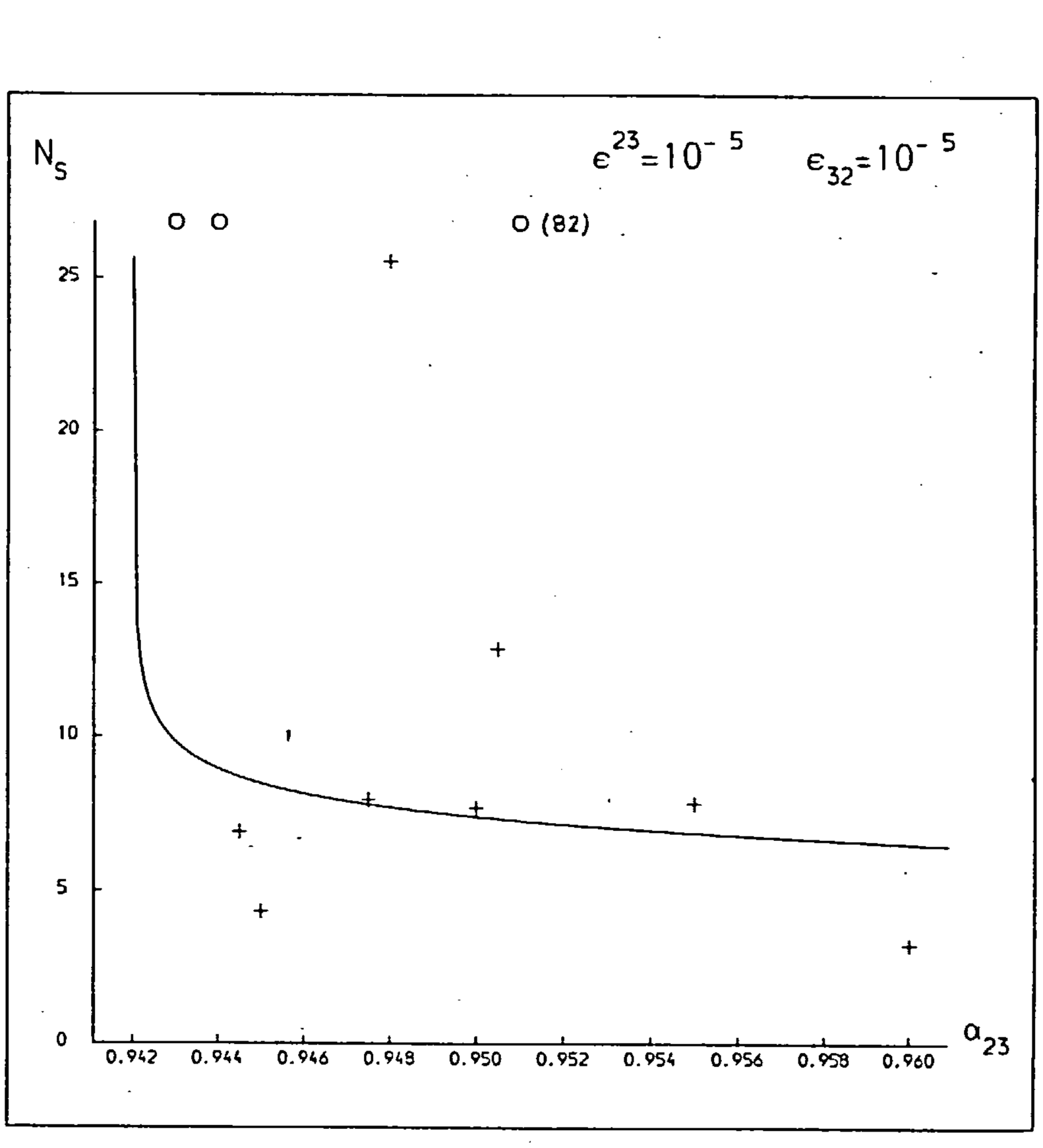

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-5}$
80
70
60
50
40
30
20
10
0
0.916
0.920
0.924
0.928
0.932
0.936
α_{23}

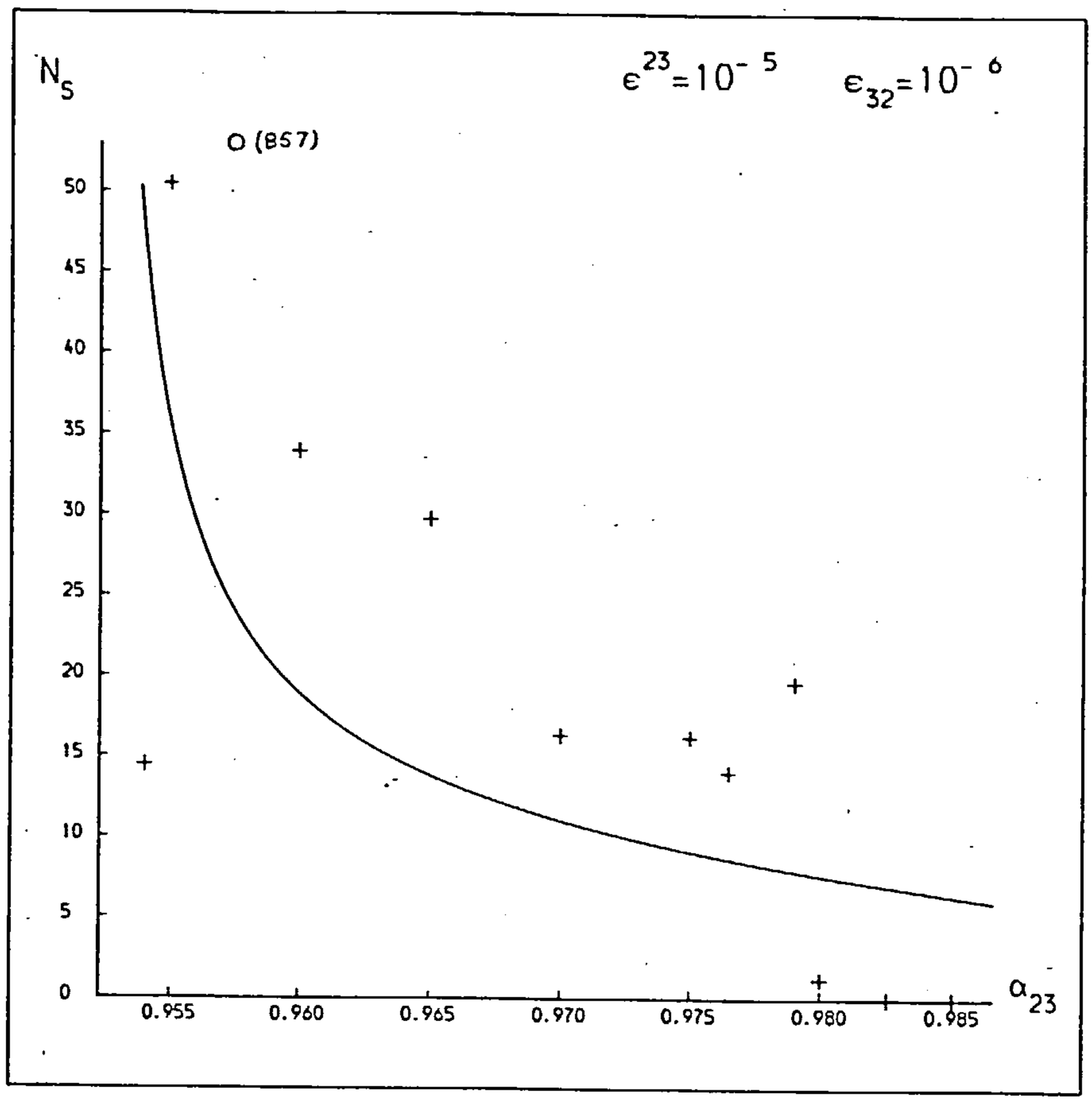

N_S
$\epsilon^{23}=10^{-4}$
$\epsilon_{32}=10^{-6}$
+ (950)
200
160
120
80
40
0
0.91
0.92
0.93
0.94
0.95
0.96
0.97
α_{23}

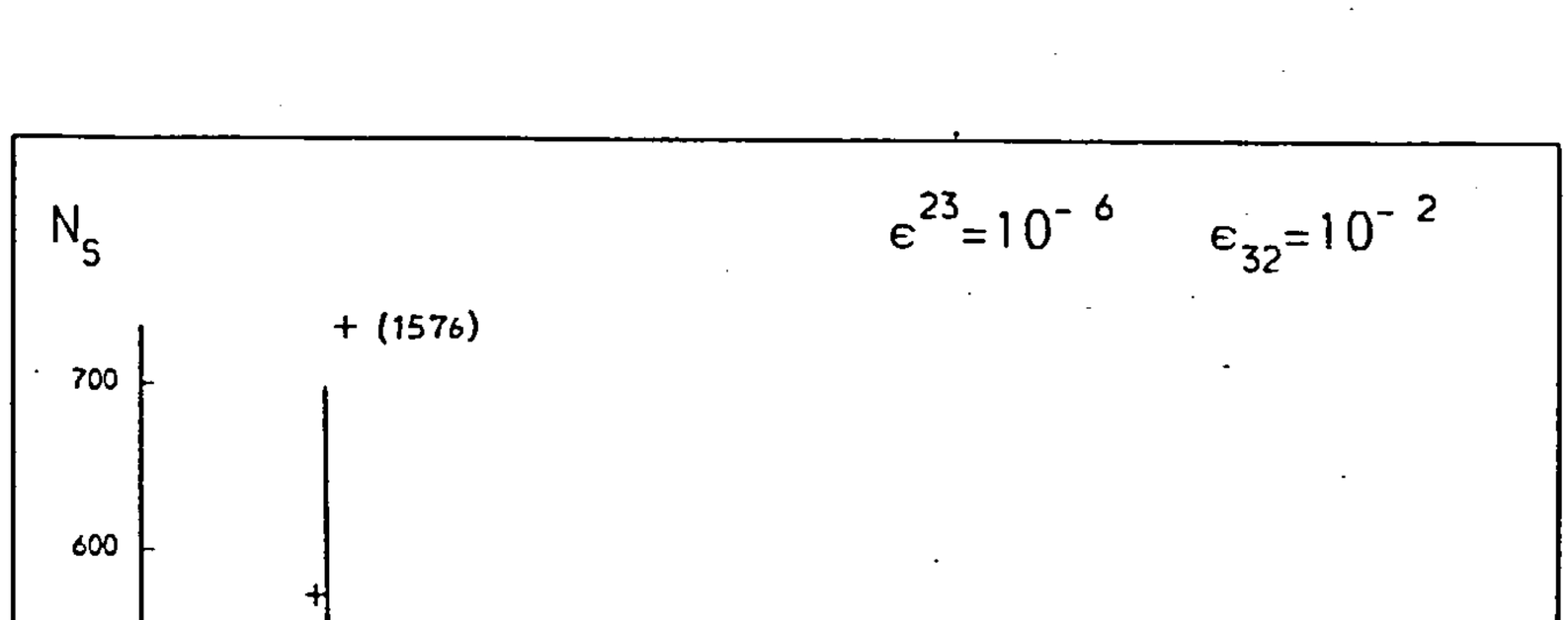

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-2}$
400
350
300
250
200
150
100
50
0
0.525
0.530
0.535
0.540
0.545
0.550
0.555
α_{23}

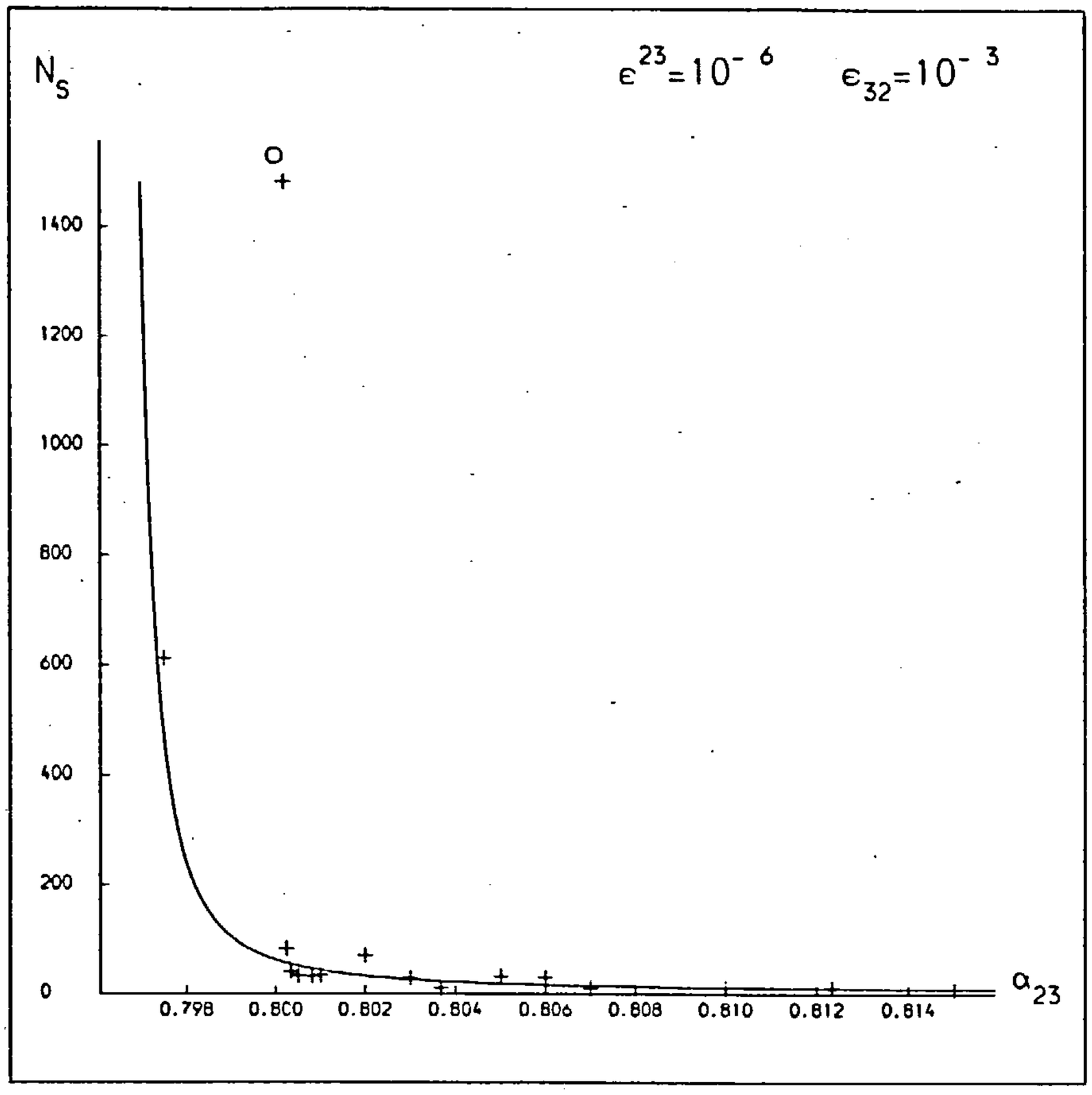

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-3}$
700
600
500
400
300
200
100
0
0.80
0.81
0.82
0.83
0.84
0.85
α_{23}

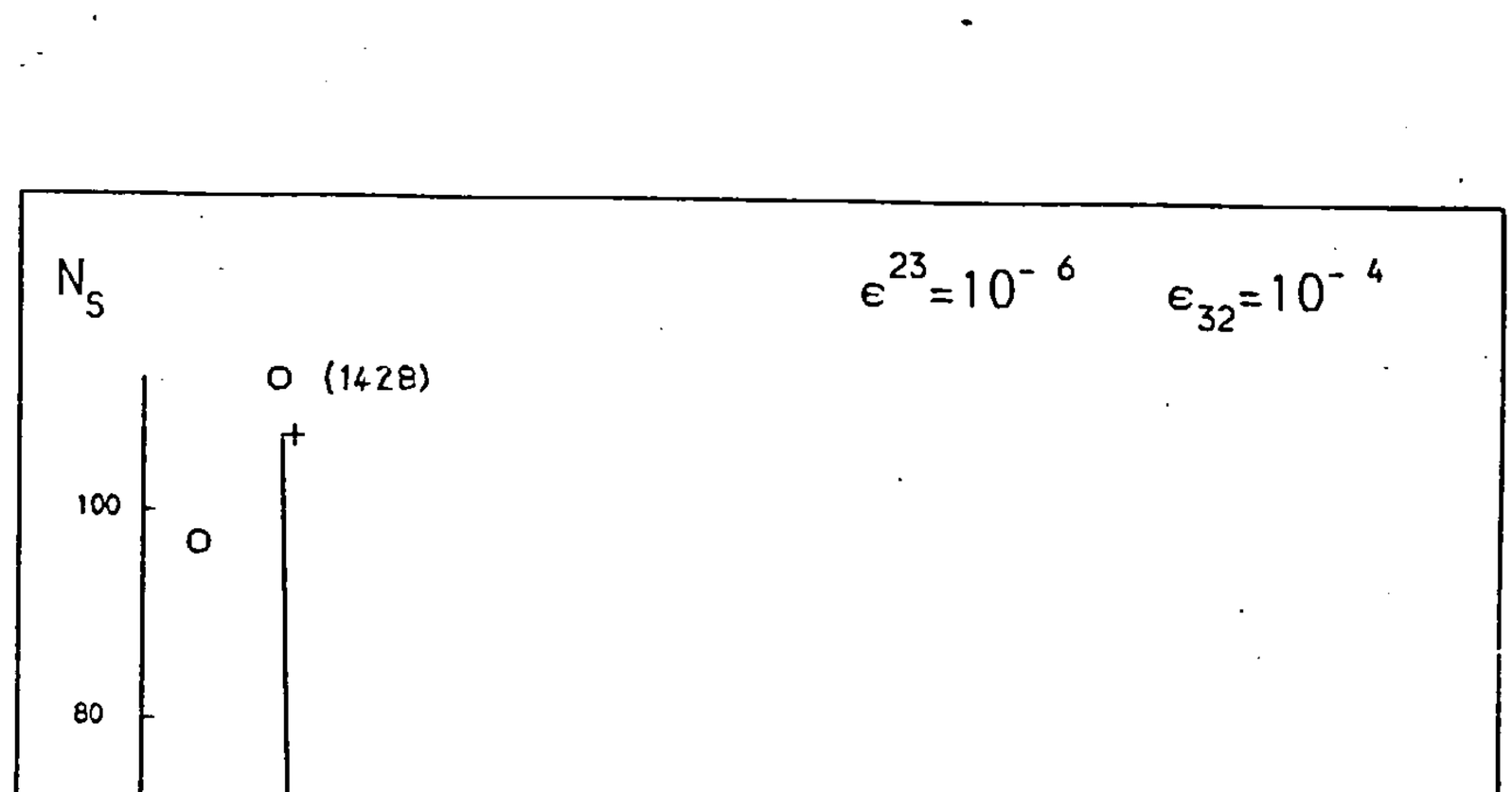

N_S
$\epsilon^{23}=10^{-5}$ $\epsilon_{32}=10^{-4}$
O (1142)
300
250
200
150
100
O
50
0
0.920
0.925
0.930
0.935
0.940
0.945
0.950
0.955
α_{23}

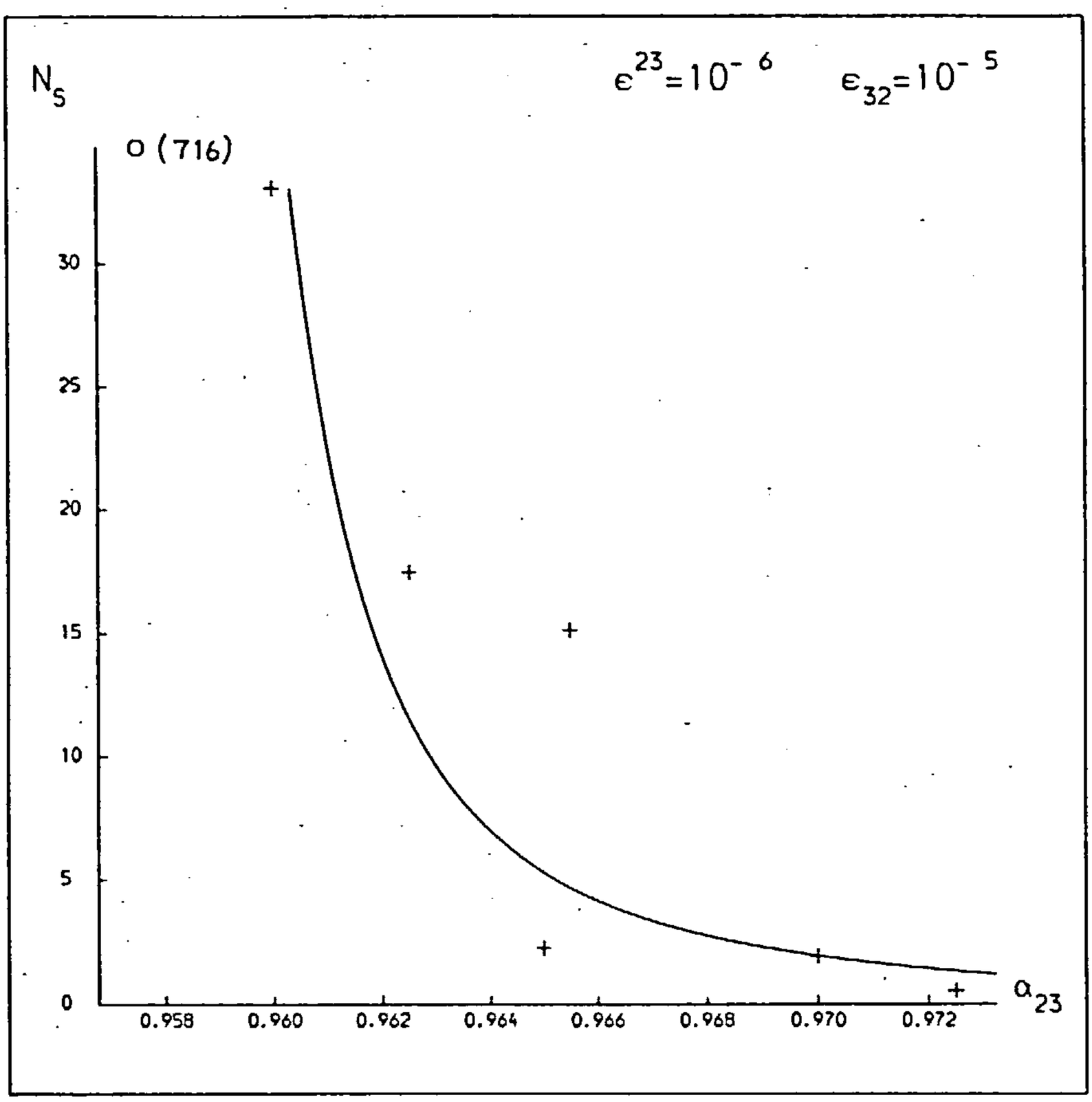

N_S
$\epsilon^{23}=10^{-5}$ $\epsilon_{32}=10^{-5}$
O O
O (82)
25
20
15
10
5
0
0.942
0.944
0.946
0.948
0.950
0.952
0.954
0.956
0.958
0.960
α_{23}

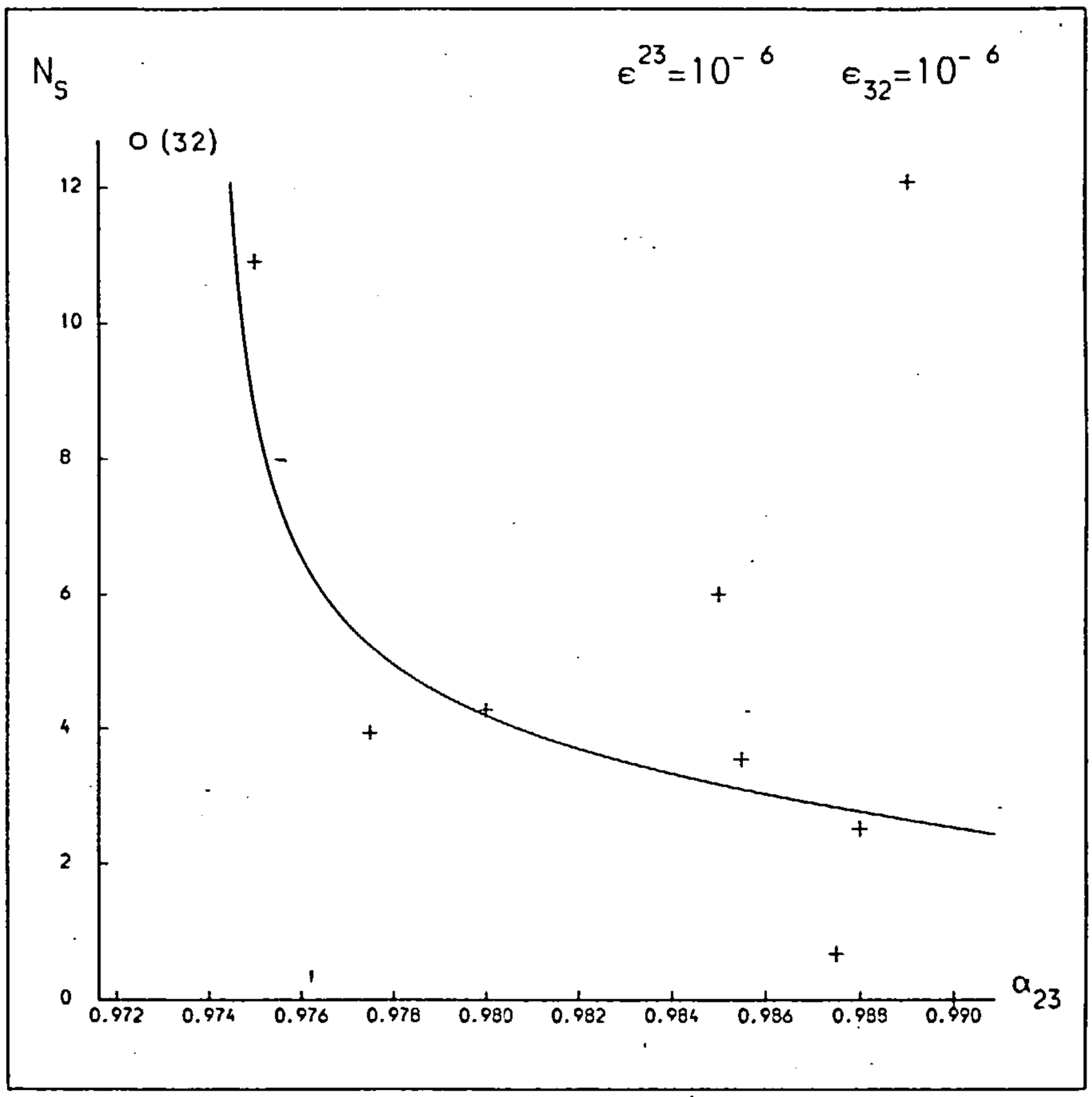

N_S
$\epsilon^{23}=10^{-5}$
$\epsilon_{32}=10^{-6}$
O (857)
50
45
40
35
30
25
20
15
10
5
0
0.955
0.960
0.965
0.970
0.975
0.980
0.985
α_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-2}$
+ (1576)
700
600
500
400
300
200
100
0
0.50
0.52
0.54
0.56
0.58
0.60
0.62
0.64
0.66
0.68
0.70
α_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-3}$
O
+
1400
1200
1000
800
600
400
200
0
0.798
0.800
0.802
0.804
0.806
0.808
0.810
0.812
0.814
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-4}$
O (1428)
O
+
100
80
60
40
20
0
0.915
0.920
0.925
0.930
0.935
0.940
0.945
0.950
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-5}$
O (716)
30
25
20
15
10
5
0
0.958
0.960
0.962
0.964
0.966
0.968
0.970
0.972
a_{23}


N_S
$\epsilon^{23}=10^{-6}$
$\epsilon_{32}=10^{-6}$
O (32)
12
10
8
6
4
2
0
0.972
0.974
0.976
0.978
0.980
0.982
0.984
0.986
0.988
0.990
a_{23}

5.2

for introducing the results of the 450 numerical experiments in the form of 25 graphs of stability lifetime N_s against α for various combinations of ϵ^{23} and ϵ_{32} (Figures 5.1). The crosses indicate the stability lifetimes of unstable systems. The circles indicate the time over which a system has been studied which as yet has shown no instability. A best fit curve is drawn through the data on each graph. The method of fitting is described in the next section.

5.3 Curve-Fitting Techniques Applied to Direct Systems.

In comparing the graphs of N_s against α for direct systems (Figures 5.1) with the graphs for retrograde systems (Figure 4.1), the most obvious difference is the effect of commensurabilities (see Sections 5.4 and 6.3). For direct systems, the commensurabilities cause a spread in the data points which was missing in the retrograde cases. This allows considerable tolerances when fitting curves to the data.

Walker and Roy, in Paper III, considered the commensurabilities as acting to enhance the stability of systems in such a way as to increase the system's lifetime beyond its "natural" value. The amount of this increase would depend on the system's proximity to a natural commensurable configuration and the inherent strength of the commensurability. With this in mind they fitted curves along the observed lower bound of the data. Some of the points would lie below the line, but the majority would lie above it. While there may be some justification for this approach on physical grounds (Section 6.3), there are problems in curve-fitting.

The first is that the "natural" lower bound is hard to plot

5.3

accurately. If the data is sparse, there may be too few points to mark it accurately. At high α values there may be so many commensurable values as to make it impossible to find. There may be damaging commensurabilities which cause a reduction in the lifetime from the "natural" value (Section 5.4). All these factors make finding the "natural" curve difficult.

The second problem is the fact that this curve is only found by eye and is open to subjective bias in deciding what really is the lower bound.

The third problem will arise when we come to combine the data from all epsilon pairs by a suitable normalisation. Unless we are able to fit the curves consistently accurately there may be a greater measure of deviation than is necessary (see Section 5.5).

The solution to these problems is to proceed as for the retrograde systems and fit a curve through the "most likely" value of the lifetime for a particular α. In other words, we fit a curve through the middle of the (α,N_s) distribution rather than the lower bound. By fitting a curve that maximises the correlation coefficient r, for all α_o,β,γ we achieve this end. The assessment of the curve is done in an objective fashion which gives the best fit available for the data.

We now summarize the curve fitting procedure for direct systems, making comparisons with retrograde systems where necessary.

For most ϵ^{23}, ϵ_{32} pairs in the retrograde case, N_s decreased monotonically as α increased and showed strong quantisation around integer numbers of synodic periods. This was manifested by the step-function nature of the data (Figures 4.1). In fitting the curves, only one representative (α,N_s) point from each "step" was used. There is no

5.3

stepping in the direct case and so all the data is used.

Because the data is not monotonic, it is not possible to assume that all systems with α less than a prescribed value will undergo, say, 600 synodic periods without exhibiting instability. That prescribed value cannot be found to any degree of accuracy and cannot therefore be used to test if a particular curve fits the "stable"* data (the circled systems in Figures 5.1). In fitting the curves, we are unable to use the data from the "stable" systems in any way. Our choice of α_o is only limited by α_c, the critical value for hierarchical preservation, and by α_{us}, the lowest value of α for which an unstable system has been detected. α_o is assumed to lie in range (α_c, α_{us}).

With these limitations, the curve fitting proceeds as for the retrograde systems. We are fitting the curve

$$N_s = f(\alpha;\alpha_o,\beta,\gamma) = \exp\left[\beta\left(\frac{1-\alpha}{\alpha-\alpha_o}\right)^{\gamma}\right]-1 , \quad \alpha_o < \alpha < 1. \qquad (1)$$

for given ϵ^{23} , ϵ_{32} . The curve is parameterised by α_o,β,γ . For a particular value of α_o, the best fit values of β and γ are found by calculating u and v for each data point (α,N_s) (Equations (4.4) and (4.5)) and constructing the least squares fit. The goodness of this fit is found by calculating the correlation coefficient r , (Equation (4.6)). The value of α_o is varied in the range (α_c,α_{us}). The best fitting values of α_o,β,γ will maximise r.

As with the retrograde systems, it is more important to get an accurate fit at the low α-high N_s end of the data than the high α-low

*A "stable" system is taken to mean one that has shown no instability during the time it has been investigated.

5.3

N_s data. To this end, the (u,v) data is weighted accordingly when performing the least squares fit to determine β and γ for a particular α_o. Care is taken to make the weighting factor vary with u, rather than v. The reason for this is that u is a function of α, (u increases as α decreases), while v is a function of N_s (v increases as N_s increases). If we were to apply weights according to how high N_s is for a particular data point, then we would weight in favour of commensurable systems as well as systems near the threshold of stability. We avoid this pitfall by weighting according to u (and hence α).

A typical graph of v against u for direct systems is given in Figure 5.2. It is clear that there will be a considerable error in the estimation of β and γ due to uncertainties in the lifetime caused by commensurabilities.

Curves are fitted in all 25 graphs in Figures 5.1, by these techniques. The critical values of α and the curve fit parameters are given in Table 5.1. It is seen from the table that α_o tends to 1 as ϵ^{23}, ϵ_{32} tend to zero. The rise is monotonic with few exceptions, one being when $\epsilon^{23} = 10^{-3}$ and $\epsilon_{32} = 10^{-3}, 10^{-4}$. It may be that the true value of α_o when $\epsilon^{23} = \epsilon_{32} = 10^{-3}$ is actually lower than stated here More numerical experiments at $\alpha < \alpha_{us}$ would be needed to verify this.

There is little evidence of any trends in β or γ as $-\epsilon^{23}$, ϵ_{32} vary. Although for $\epsilon^{23} = 10^{-2}$, β increases and γ decreases as ϵ_{32} decreases, there is no evidence from $\epsilon^{23} < 10^{-2}$ that this is anything other than chance.

For most curves, the agreement is reasonable. 16 out of 25 show $r > 0.8$. Occasionally a fit is less than satisfactory ($r < 0.5$) and in some cases $r < 0$ indicating a slight anti-correlation. Generally

5.3

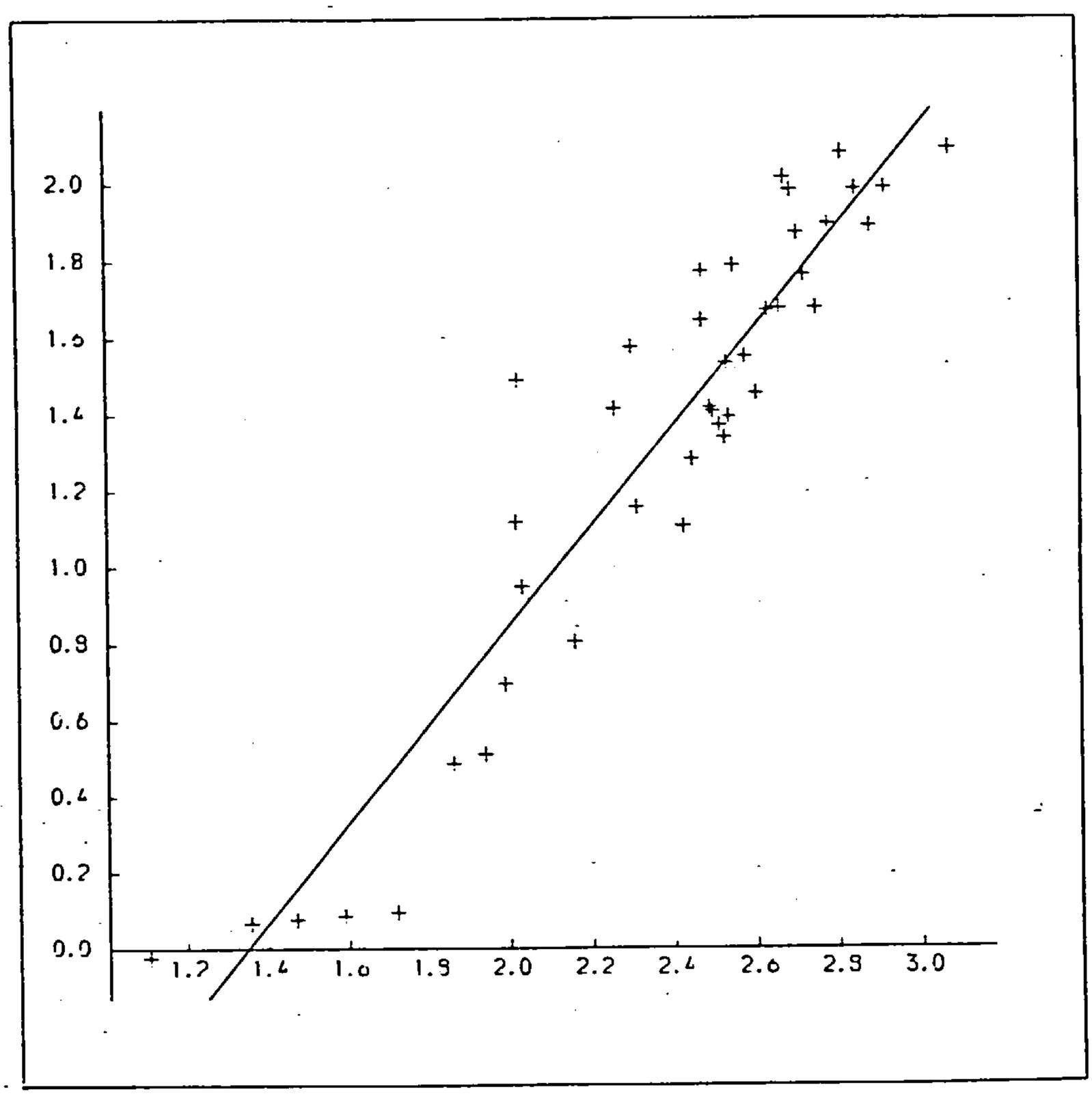


Figure 5.2: Graph of v against u for $\varepsilon^{23} = \varepsilon_{32} = 10^{-2}$ (direct systems).

there are two reasons for this. When $r \prec 0$, there are usually commensurable systems at high α with anomalously long lifetimes. When these points are excluded the curve fits well to the rest of the data and $r > 0.8$.

ϵ^{23}	ϵ_{32}	α_c	α_o	α_{us}	β	γ	r
10^{-2}	10^{-2}	0.324	0.434	0.459	0.174	1.298	0.94
	10^{-3}	0.495	0.598	0.620	0.449	0.939	0.89
	10^{-4}	0.518	0.620	0.625	0.506	0.595	0.91
	10^{-5}	0.520	0.632	0.633	0.920	0.309	0.81
	10^{-6}	0.521	0.627	0.628	1.156	0.268	0.85
10^{-3}	10^{-2}	0.397	0.510	0.511	0.739	0.319	0.94
	10^{-3}	0.707	0.792	0.793	0.472	0.479	0.86
	10^{-4}	0.768	0.783	0.788	0.811	0.539	0.87
	10^{-5}	0.776	0.794	0.795	0.604	0.397	0.75
	10^{-6}	0.777	0.795	0.796	0.687	0.324	0.87
10^{-4}	10^{-2}	0.412	0.519	0.520	1.017	0.284	0.77
	10^{-3}	0.760	0.786	0.790	0.120	0.996	0.93
	10^{-4}	0.871	0.902	0.905	0.782	0.495	0.86
	10^{-5}	0.896	0.915	0.917	0.675	0.384	0.56
	10^{-6}	0.899	0.914	0.915	0.110	0.890	0
10^{-5}	10^{-2}	0.413	0.492	0.525	0.062	1.703	0.83
	10^{-3}	0.767	0.798	0.799	0.965	0.303	0.90
	10^{-4}	0.894	0.917	0.918	0.260	0.640	0
	10^{-5}	0.942	0.942	0.945	1.940	0.051	0.16
	10^{-6}	0.952	0.952	0.954	2.283	0.169	0.46 *
10^{-6}	10^{-2}	0.413	0.518	0.520	0.915	0.401	0.68
	10^{-3}	0.767	0.796	0.798	0.885	0.395	0.89
	10^{-4}	0.897	0.916	0.918	0.451	0.596	0.82
	10^{-5}	0.953	0.953	0.958	0.598	1.050	0.85
	10^{-6}	0.973	0.974	0.975	1.360	0.160	0.39

Table 5.1: Summary of critical values of α obtained from the results displayed in Figures 5.1. Curve fit parameters and correlation coefficients are also given. These results apply for direct systems only.

*These values are derived from an unweighted fit. If the weighted fit is used, $\beta = 4.336$, $\gamma = -0.109$, $r = 0$.

5.3

The second reason is that the data seems genuinely scattered. This is generally the case when both ϵ^{23} and ϵ_{32} are small. Walker and Roy experienced similar difficulties, quoting the high number of "one-spoke" commensurabilities as being responsible (see Section 5.4). In these cases the values of α_o, β, γ should be treated with some scepticism.

5.4 The Effect of Commensurabilities.

It has been seen in Figure 5.1 that commensurabilities seem to play an important part in determining the stability lifetimes of direct three-body systems. It is interesting to note which commensurabilities may be responsible in individual cases.

Let n_2 and n_3 be the mean motions of the (m_1, m_2) and (M_2, m_3) binary systems respectively $(n_2 > n_3)$, these being the osculating values, given by

$$n_2^2 \; a_2^3 = G(m_1 + m_2)$$

$$n_3^2 \; a_3^3 = G(m_1 + m_2 + m_3)$$

A system is defined to be commensurable when the ratio of mean motions approximates to a ratio of integers, i.e.

$$\frac{n_2}{n_3} = \frac{A_2}{A_3} \quad , \quad A_2, A_3 \in \mathbb{Z}, \; A_2 > A_3$$

Dividing the first equation by the second, we get

$$\alpha^3 = \left(\frac{n_3}{n_2}\right)^2 \frac{1}{1+\mu_3}$$

where $\mu_3 = m_3/(m_1+m_2)$ as before. Recalling that $\epsilon_{32} = \mu_3 \alpha^3$,

5.4

$$\alpha = \left[\left(\frac{A_3}{A_2}\right)^2 - \varepsilon_{32}\right]^{1/3}$$

Hence we find the value α which gives rise to the $A_2:A_3$ commensurability. Note that the equation is independent of ε^{23}. In addition, it depends only weakly on the value of ε_{32}, when ε_{32} is small.

This allows us to give a simple table of the range of values of α for the strongest commensurabilities (Table 5.2). For each pair A_2, A_3, two values of α are given. The lower corresponds to the value at $\varepsilon_{32} = 10^{-2}$; the higher at $\varepsilon_{32} = 10^{-6}$. For $\varepsilon_{32} = 10^{-3}$, 10^{-4}, 10^{-5}, the values of α lie closer to the higher value than the lower value. Obviously the 6:3 commensurability is equivalent to the 2:1 commensurability. The dashes in Table 5.2 indicate commensurabilities that are equivalent to others already given.

Table 5.3 gives for each of the graphs in Figures 5.1, the values of α at which commensurable behaviour is detected from the numerical integration experiments. Along side are presented the commensurabilities that are suspected of being responsible and the corresponding exact value of α. This follows Walker and Roy's treatment, Paper III. They noted that commencing each numerical experiment, the osculating value of α was seen to decrease. It therefore seems likely that any commensurability will only affect systems whose initial value of α is greater than the value associated with the commensurability.

The osculating synodic period of a system $S = 2\pi/(n_2 - n_3)$. From Equation (4),

$$S = \frac{2\pi}{n_2} \cdot \frac{A_2}{A_2 - A_3}$$

A_2						
2	0.621 0.630					
3	0.466 0.481	0.757 0.763				
4	0.374 0.397	- -	0.821 0.825			
5	0.311 0.342	0.531 0.543	0.705 0.711	0.857 0.862		
6	0.261 0.303	- -	- -	- -	0.881 0.886	
7	0.218 0.273	0.415 0.434	0.558 0.568	0.682 0.689	0.794 0.799	0.898 0.902
	1	2	3	4	5	6 A_3

Table 5.2: The values of α which give rise to commensurabilities $A_2:A_3$. for $10^{-6} \lesssim \varepsilon_{32} \lesssim 10^{-2}$. (The values are independent of ε^{23}). The two entries at each A_2,A_3 are the minimum and maximum values of α for ε_{32} in this range.

Table 5.3 : Observed commensurabilities in mean motion for direct three-body systems.

ε^{23}	ε_{32}	α	Commensurability/Corresponding α	
10^{-2}	10^{-2}	0.466)		
		0.47)		
		0.475)	3:1	0.466
		0.485)		
		0.500	11:4	0.496
10^{-2}	10^{-3}	0.635	2:1	0.629
10^{-2}	10^{-4}	0.630	2:1	0.630
10^{-2}	10^{-5}	0.630	2:1	0.630
10^{-2}	10^{-6}	0.670	11:6	0.668
		0.630	2:1	0.630
10^{-3}	10^{-2}	-	-	-
10^{-3}	10^{-3}	0.795	7:5	0.795
10^{-3}	10^{-4}	0.792	17:12	0.7925
		0.805	18:13	0.805
		0.822	4:3	0.822
10^{-3}	10^{-5}	0.818	11:8	0.809
		0.85	9:7	0.846
10^{-3}	10^{-6}	0.798)	7:5	0.799
		0.802)		
		0.818)		
		0.820)	4:3	0.825
		0.822)		
10^{-4}	10^{-2}	0.532	5:2	0.531
		0.546	12:5	0.546
10^{-4}	10^{-3}	-	-	-
10^{-4}	10^{-4}	0.908	15:13	0.909
		0.95	13:12	0.948
10^{-4}	10^{-5}	0.920	8:7	0.915
		0.927	9:8	0.925
		0.936	10:9	0.932
10^{-4}	10^{-6}	0.95	13:12	0.948

Table 5.3: continued...

ϵ^{23}	ϵ_{32}	α	Commensurability/Corresponding α.	
10^{-5}	10^{-2}	0.533	5:2	0.531
		0.545	12:5	0.547
10^{-5}	10^{-3}	0.800	7:5	0.799
		0.808	11:8	0.808
		0.83	4:3	0.825
10^{-5}	10^{-4}	0.94	11:10	0.938
		0.955	15:14	0.955
10^{-5}	10^{-5}	0.951	14:13	0.952
10^{-5}	10^{-6}	-	-	-
10^{-6}	10^{-2}	0.525	-	-
		0.530)	5:2	0.531
		0.535)		
		0.57)	9:4	0.572
		0.60)		
10^{-6}	10^{-3}	0.800)	7:5	0.799
		0.802)		
		0.805	18:13	0.804
10^{-6}	10^{-4}	0.925	9:8	0.924
		0.940	11:10	0.938
10^{-6}	10^{-5}	-	-	-
10^{-6}	10^{-6}	0.989	-	-

5.4

In this form, it can be seen that straight line configurations can only occur along a finite number $B = A_2 - A_3$ of sidereal directions. The system of conjunction lines can be pictured like the spokes of a wheel. The wheel has B spokes with each successive conjunction taking place A_3 spokes away from the previous spoke. For example, let $n_2:n_3 = 5:2$ as in Figure 5.3.

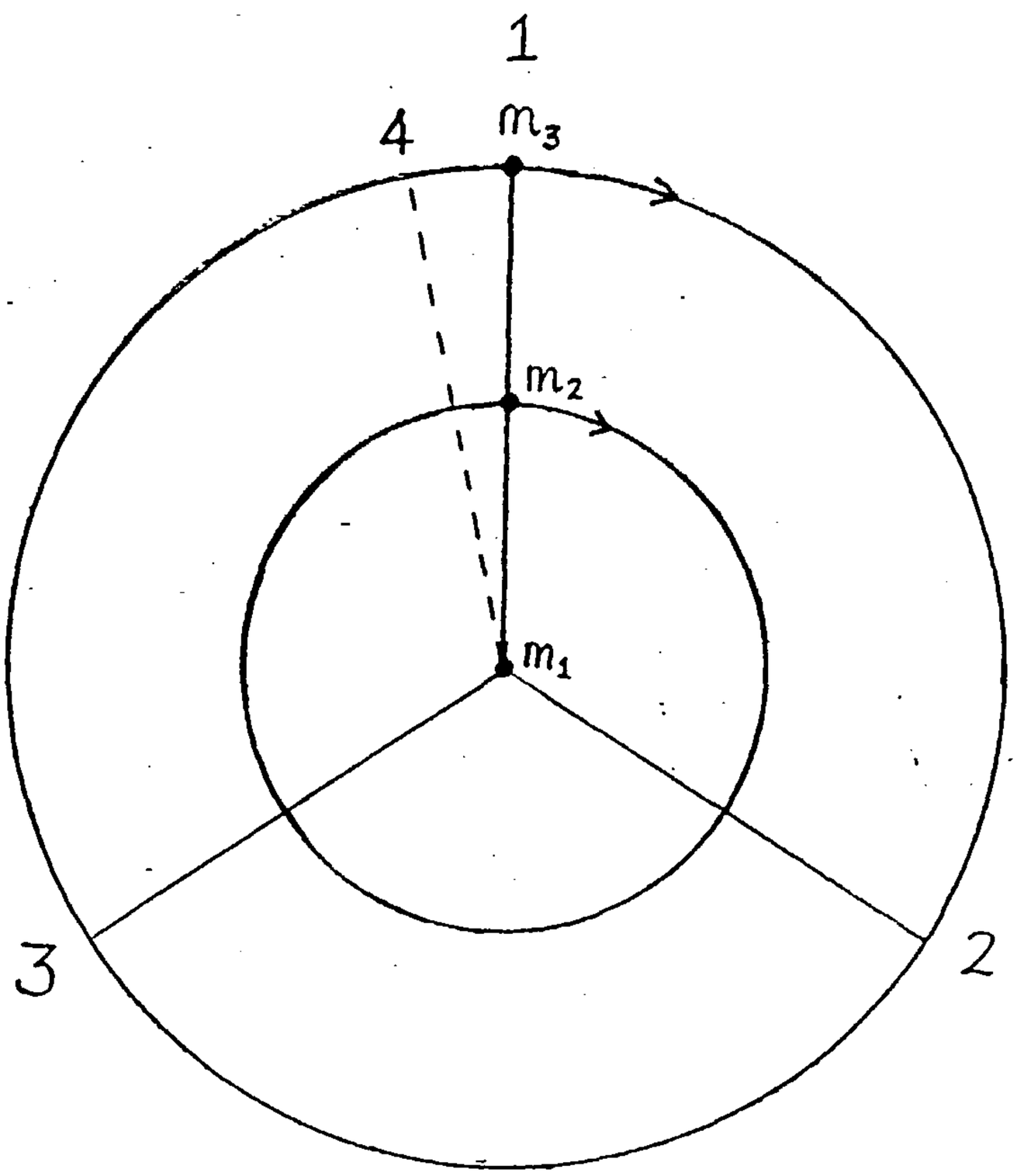


Figure 5.3: System in 5:2 commensurability will be at conjunction in directions 1-3-2 consecutively. If the commensurability is exact the conjunction lines will remain fixed in positions 1,2,3. If not the fourth conjunction line will be displaced from the first by a small amount.

5.4

There are 3 spokes numbered clockwise (in the direction of rotation of the bodies). If the system starts at conjunction in direction 1, after one synodic period $S = 5/3 \times 2\pi/n_2$, there will be a conjunction in direction 3. After a further synodic period, the conjunction will take place in direction 2, and so on. A 4:1 commensurable system would also have three spokes but they would be visited in the order 1,2,3. If the commensurability is exact then spokes are fixed in the same sidereal directions. If the commensurability is inexact, the sytem of conjunction lines will rotate, the speed of rotation depending on how inexact the commensurability is.

We have already seen that the system is generally at its most unstable near conjunction, where the perturbations are greatest. In particular, the perturbations will be greatest when the conjunction takes place at the apocentre of the (m_1,m_2) binary and the pericentre of the (M_2,m_3) binary.

A commensurability limits the number of sidereal directions where conjunctions are allowed to take place. It may be that such a commensurability prevents a system from approaching this worst possible conjunction which may cause an instability to appear. If the commensurability is inexact then there is only a temporary reprieve as the conjunction lines move towards the (m_1,m_2) apocentre - (M_2,m_3) pericentre configuration. Therefore the more exact the commensurability, the more likely the system is to survive for anomalously long periods of time.

Clearly commensurabilities with few distinct conjunction lines are more likely to preserve stability in this way than commensurabilities showing many conjunction lines, in order to minimise the number of

5.4

conjunction positions. Thus the commensurabilities close to the diagonal in Table 5.2 are favoured. In particular commensurabilities of the form n + 1 : n have only one conjunction line and are the most favourable. Such commensurabilities occur for high α (> 0.9) and it may be that this explains the very random nature of the lifetimes shown when both ϵ^{23} and ϵ_{32} are less than or equal to 10^{-5}.

Table 5.4 gives the frequency of B arising from the commensurabilities associated with the systems which exhibited anomalously long stability lifetimes. It does indeed seem that "one spoke wheels" are favoured. It also appears that odd values of B are favoured over even values.

This should not be surprising since for even valued B, if one commensurability was well placed at (m_1, m_2) pericentre - (M_2, m_3) apocentre,

Table 5.4: Frequency of observed numbers of conjunction lines.

B	1	2	3	4	5	6	7
Frequency	17	7	5	0	5	0	3

then there would be another conjunction at (m_1, m_2) apocentre - (M_2, m_3) pericentre. Thus best and worst conjunctions always go together. It is however worth testing the significance of this data to see if in reality, even B are avoided.

The probability that B is even is 1/3.

5.4

Sketch Proof

It is assumed that in any quoted commensurability the greatest common divisor of A_2, A_3 is one. To begin however, we pick integers A_2 and A_3 at random, such that $A_2 > A_3$. Each has a 50% chance of being odd or even. Thus there are four cases to be considered, namely, (A_2 odd, A_3 odd), (A_2 even, A_3 odd), (A_2 odd, A_3 even), (A_2 even, A_3 even).

By considering A_2 and A_3 as products of prime numbers it is easily seen that any common divisor of A_2 and A_3 is odd, if one of them is odd. Thus in the first three cases when the ratio $A_2:A_3$ is reduced so that the greatest common divisor is one, the parities of A_2 and A_3 are preserved.

In the case when A_2 and A_3 are both even, we may divide both by 2 until one or both are odd in which case they fall into one of the first three categories, each of which is equally likely to occur.

When A_2 and A_3 are both odd then B is even. Otherwise, one is odd and the other is even, implying B is odd. Thus the probability of B being even is $^1/3$.

Q.E.D.

The probability of r even spoked commensurabilities in n trials is given by the binomial distribution B(n,p) where p is the probability of an even spoked commensurability in one trial. In this case $p = {}^1/3$, $n = 37$. We use the normal approximation to the binomial distribution, namely $B(n,p) \simeq N(np,npq)$. Thus $B(37, {}^1/3) \simeq N(12.33, 8.22)$, $(q=1-p)$. We detected 7 even spoked commensurabilities out of 37. The probability of having as few as that is

5.4

$$P(r \leq 7) = \Phi\left[\frac{7.5 - 12.33}{\sqrt{8.22}}\right]$$

$$= \Phi(-1.68)$$

$$= 1 - \Phi(1.68)$$

$$= 1 - 0.9535$$

$$= 0.0465$$

where Φ is the cumulative distribution function of N(0,1). Thus the probability of seeing as few even spoked commensurabilities is less than 5% and to this level of significance we must assume that systems do avoid such commensurabilities.

Recall in Chapter 4 that for retrograde systems, there was an anomalous peak in the (α, N_s) distribution for some values of ε^{23} , ε_{32} . There is also some evidence of similar peaks for direct systems. In the cases $(\varepsilon^{23}, \varepsilon_{32}, \alpha) = (10^{-3}, 10^{-4}, 0.805)$, $(10^{-3}, 10^{-5}, 0.815)$, $(10^{-3}, 10^{-6}, 0.82)$, $(10^{-6}, 10^{-2}, 0.57)$ there is evidence of a broad feature, (see Figures 5.1 and Table 5.3). Although there is a strong commensurability in the neighbourhood of each point, the strength of the feature is much greater than other commensurabilities. It is this that reminds us of similar peaks for retrograde systems.

Walker has pointed out that commensurable behaviour can arise when a system is not initially commensurable because the osculating value of α decreases immediately after the start of the integration. It may be that as the system changes in this manner, it becomes commensurable for a time and this commensurability serves to enhance the overall stability. It is however possible that the accumulated error in the numerical procedure could enhance the stability in the same way. The method of

5.4

rounding by the computer as well as the truncation error, may introduce a secular trend to a_2 and a_3 which will cause a dissipative effect in what should be a conservative system. Goldreich (1965) suggests that dissipation allows systems to evolve into commensurable ones. Obviously, if the numerical error is small, this effect will be small compared to the real dynamical effects that are taking place. If however we run a commensurable system for a long time in the hope of detecting an instability then we may be running it in vain; not because it is indeed stable, but because the numerical procedure renders it so. It is believed however that in 4000 synodic periods (150,000 integration steps) the numerical errors are still small enough that this effect is neglible.

A further cautionary note is perceived, arising from the real problem of deciding which commensurability is "guilty of inciting stability" for a system that shows an anomalously long stability lifetime. For example, the system may lie reasonably close to a strong commensurability like the 5:3, with only 2 conjunction lines that are rotating fairly quickly. On the other hand it may be equally close to the 26:15 commensurability with 11 conjunction lines that are almost stationary. Which commensurability has the greater effect: an inexact "strong" commensurability or an exact "weak" commensurability? One can go further. Because the set of rational numbers is dense on the real line, we may pick for any system, two integers A_2, A_3 such that n_2/n_3 is arbitrarily close to A_2/A_3.

This leads directly on to the final point in this section. For the retrograde systems, we were fairly convinced of the value of α_o, the $N_s(\alpha)$ curve tending to a definite asymptote. This is no longer the case for the direct systems. There is always the possibility that what seems to be

5.4

an asymptotic rise to infinity as α decreases, could be another commensurability and that the value of the lifetimes may decrease again for even lower α . An example might be for $\varepsilon^{23} = 10^{-2}$, $\varepsilon_{32} \lesssim 10^{-3}$. Here α_o is close to 0.63 the value for the 2:1 commensurability. Here, however, since the rise in lifetimes takes place for $\alpha < 0.63$, we are reasonably confident that it is truly asymptotic.

Commensurabilities have been discussed at some length in this section. In particular we have noted the difficulty in determining possible commensurable behaviour from a simple analysis of the initial parameters ε^{23} , ε_{32} , α . This in turn makes it very difficult to predict the stability lifetime in advance. We can take for granted the existence of commensurabilities and treat them as noise on some "true" $N_s(\alpha)$ curve which would exist if they were not present as for the retrograde systems. This allows us to treat them in a statistical sense as detailed in the next section. By this method it is possible to derive quantitative results regarding the stability lifetimes.

5.5 Predictions of Stability Lifetimes

In attempting to fit curves to the data in Figures 4.1 and 5.1, it has been assumed that for any $(\varepsilon^{23}, \varepsilon_{32})$ pair there exists a value α_o such that $N_s(\alpha) \to \infty$ as $\alpha \to \alpha_o+$. We have also assumed that $N_s = 0$ for $\alpha \simeq 1$. The mathematical form used to model the lifetimes in the range $(\alpha_o, 1)$, given in Equation (1), was considered to be entirely adequate for the retrograde systems of Chapter 4. It is not however as accurate in considering the behaviour of direct systems.

We have seen in Section 5.4 that commensurabilities play an important part in determining the stability lifetime of a given system.

5.5

Individual commensurabilities increase or decrease the stability of a system, the result being a marked uncertainty in the overall distribution of (α, N_s) points for given ε^{23} , ε_{32} . One may presume that every direct system is affected by one or more commensurabilities. We note that the set of commensurable systems is dense on the real α-axis. Given both these conditions, it may be fruitless to search for a continuous curve that models N_s against α . One may hypothesize a stochastic or fractal nature to the function $N_s(\alpha)$, (Mandelbrot, 1982).

If this is indeed the case, a more relevant question might be, "How accurate is our estimate of the lifetime, $N_s(\alpha)$, likely to be"? We are now considering for a given system (defined as always by (ε^{23} , ε_{32} , α)) the lifetime N_s as a random variable to which we assign a probability density function (p.d.f.). If we are able to find a suitable p.d.f., $h(N_s)$ say, then by definition the probability of that system having a stability lifetime between two values $(N_s)_1$ and $(N_s)_2$ is

$$P((N_s)_1 < N_s < (N_s)_2) = \int_{(N_s)_1}^{(N_s)_2} h(N_s^*)dN_s^*$$

where $(N_s)_1$, $(N_s)_2 \geq 0$. We require that h has the following basic properties:-

1) $h(N_s) = 0 \qquad \forall N_s \leq 0$

2) $h(N_s) > 0 \qquad \forall N_s > 0$

3) $h(N_s) \to 0 \quad$ as $\quad N_s \to \infty$

4) h has a functional form that is independent of $\varepsilon^{23}, \varepsilon_{32}$, α although the parameters that characterise that form may vary with ε^{23} ; ε_{32} , α .

5.5

Property (1) arises from the definition of N_s as the stability lifetime. Properties (2) and (3) arise from the definition of a p.d.f. Property (4) should provide few problems if the form of h is characterized by sufficient parameters. $(p_i;\ i=1,\ldots,m)$.

Given these properties, $h(N_s)$ would be different for each system, since we are assuming that each p_i is a function of ϵ^{23} , ϵ_{32} , α . In order to find accurate values of the parameters a curve-fitting procedure would have to be adopted. This would require numerical integration experiments being performed on many different systems with neighbouring values of ϵ^{23} , ϵ_{32} , α. This seems a pointless exercise since the object of the procedure is to avoid lengthy numerical integrations. We would be faster and more accurate if we numerically integrated the system in question. Quite clearly the existing data sets for each ϵ^{23} , ϵ_{32} are not nearly large enough to perform reliable fitting procedures. If we are to proceed at all, we must either find an analytical form that describes $p_i = p_i\ (\epsilon^{23}\ ,\ \epsilon_{32},\ \alpha)$ or increase the data set at our disposal.

The solution is to adopt both methods in part. We assume (without proof) that there exists a variable $x = x\ (\alpha;\ \epsilon^{23},\ \epsilon_{32})$ such that the parameters p_i are functions of x alone. The chosen form of x is given by

$$\frac{1}{x} = \beta\left(\frac{1-\alpha}{\alpha-\alpha_o}\right)^{\gamma} + 1 \qquad (2)$$

where α_o, β, γ are as defined in Section 5.3. Substituting Equation (2) into Equation (1) yields

$$N_s = g(x) = e^{\frac{1}{x}-1} - 1 \qquad (0 < x \leq 1) \qquad (3)$$

5.5

Through a change of variable from α to x, via the parameters α_o, β, γ we have arrived at an equation for the stability lifetime that is independent of ϵ^{23}, ϵ_{32}. In other words we have taken the (α, N_s) data for each pair of epsilons, $\alpha_o < \alpha \leq 1$ and produced normalised data (x, N_s) that lies in the range $0 < x \leq 1$. In this way the data from each epsilon pair can be superimposed on the one graph of N_s against x with a common best fit curve, (see Figure 5.4). Note that there is no single value of N_s for a given value of x.

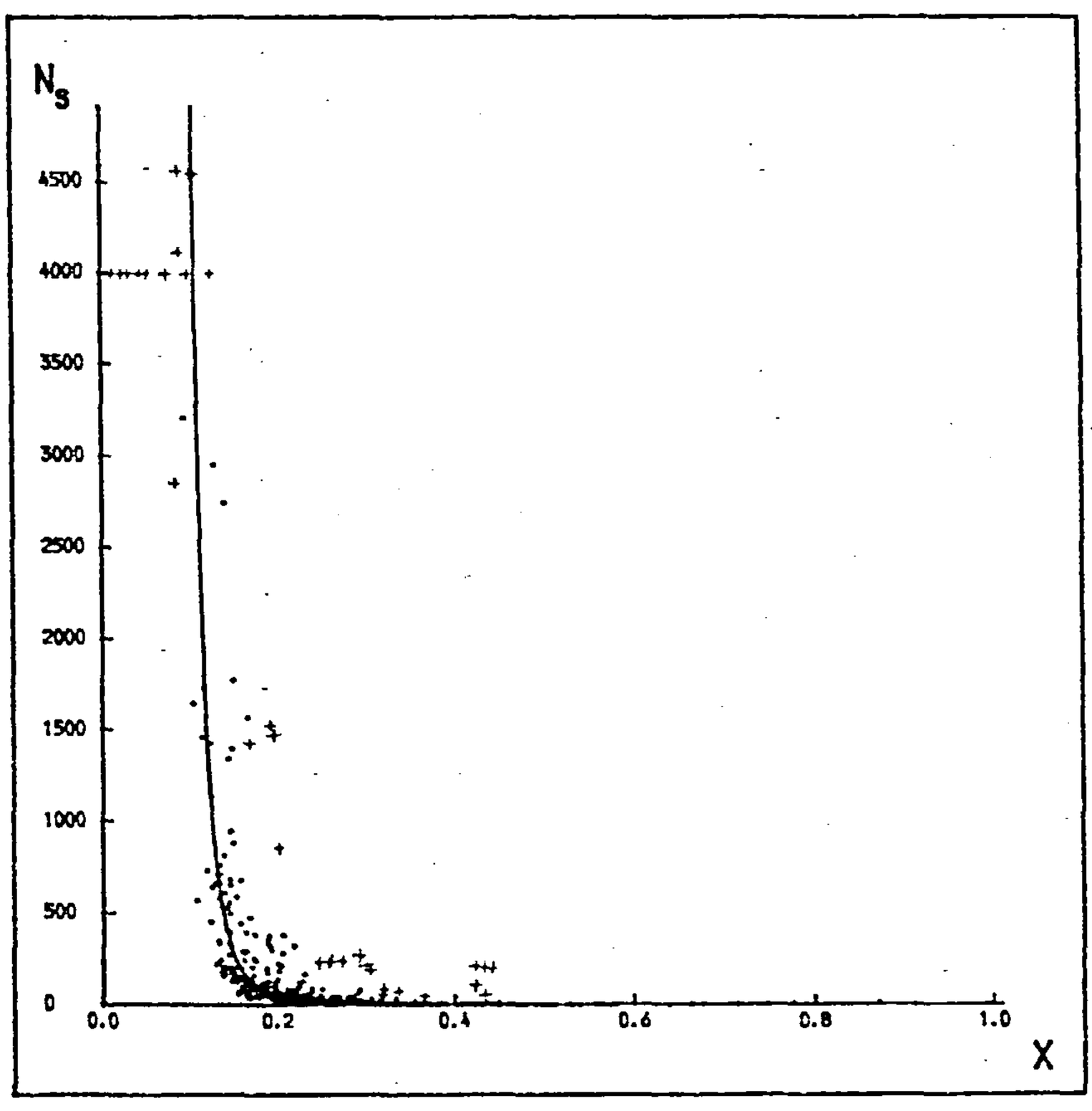


Figure 5.4: Stability lifetime against x. The unstable systems are dotted with the best fit curve drawn through them. The crosses represent anomalously stable systems, which are rigorously defined below.

5.5

Through this procedure we have simplified the determination of the parameters p_i, as they now vary only with x, and increased the data set with which we are able to carry out curve-fitting procedures. It is assumed that the variance in the lifetimes with respect to the best fit curves are comparable over all epsilon pairs. This being so, we may proceed to determine a suitable p.d.f. Three will be considered: (i) an empirically constructed p.d.f., (ii) the Gamma Distribution, (iii) the Chi-Square Distribution.

(i) Empirical Method.

For each point in the data set (x,N_s) it is possible to compute the normalised residual from the predicted value, namely

$$S = \frac{N_s - g(x)}{(g(x))^{\frac{1}{2}}} \qquad (4)$$

where g(x) is given in Equation (3). A frequency plot of S for the total data set of 450 points is given in Figure 5.5. This figure is rather encouraging since it shows a sharp peak around zero. This indicates that the majority of points do lie close to the best fit curve that has been devised. It can also be seen that the distribution is skew. This is entirely to be expected since the residuals are bounded below, but unbounded above, i.e. there is more scope for large positive values of s than for negative values.

We are however concerned with evaluating probable stability lifetimes at a particular value of x,x* say. To do this, the easiest way is to consider a sample of the data set in the neighbourhood of x*, i.e. choose $(x_i,(N_s)_i)$ where $x^* - \Delta \leq x_i \leq x^* + \Delta$, for all possible i. Order the lifetimes $(N_s)_i$ in increasing size. We therefore have a sample of life-

5.5

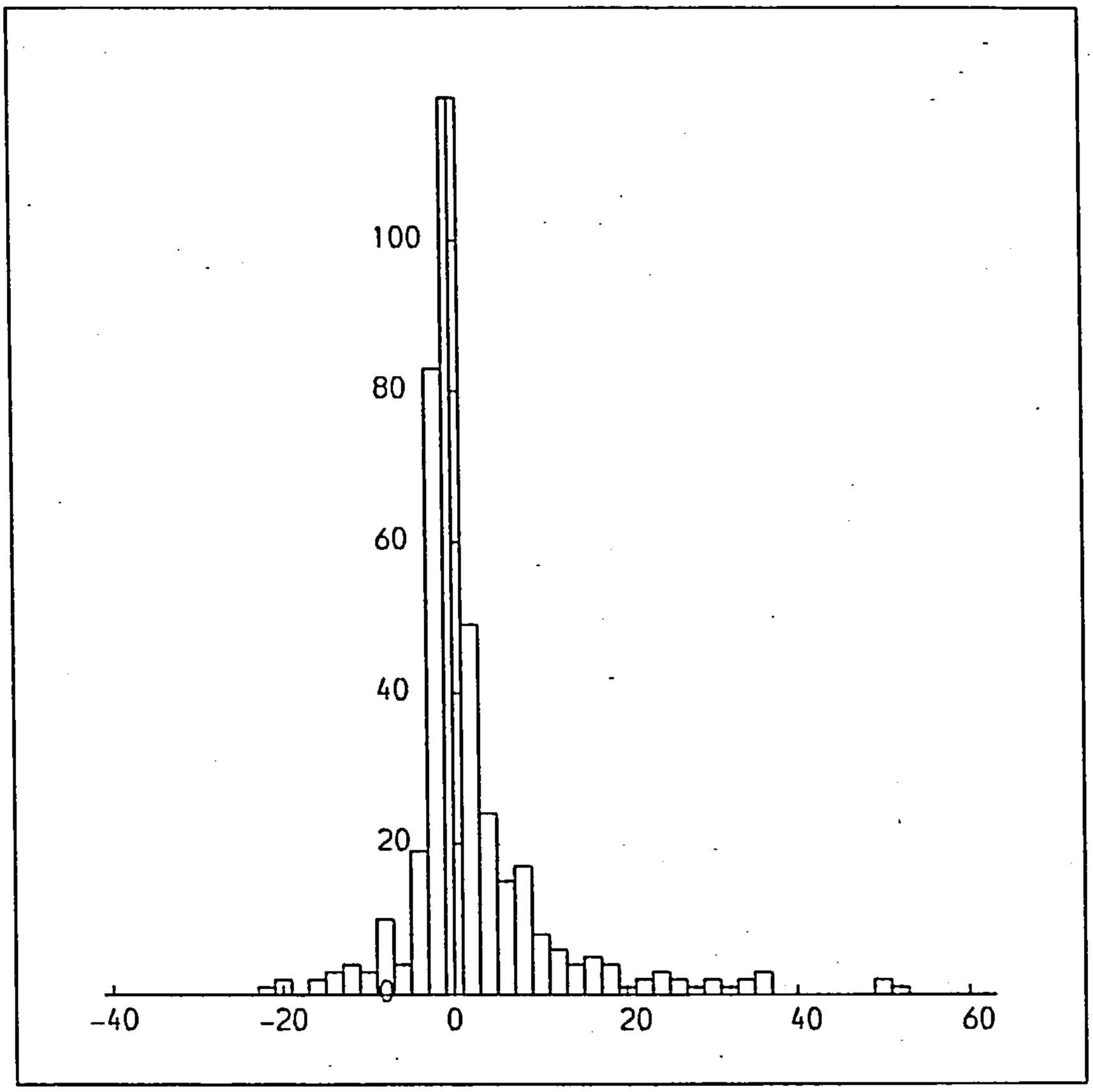


Figure 5.5: Frequency diagram of normalised residuals of stability lifetimes.

times from which we can make predictions. We do so by noting various percentile values within the sample. For example, in a sample of 50 lifetimes, we might note the 25th or 26th value as the median of the distribution. We should then say that a system characterized by an x-value equal to x* would stand a 50% chance of surviving for the time given by $(N_s)_{25}$. In a similar fashion we may note other percentiles,

5.5

such as $(N_s)_5$ (10%), $(N_s)_{10}$ (20%), $(N_s)_{40}$ (80%), $(N_s)_{45}$ (90%). From these values we can make several kinds of deductions, depending on our interests. For example, the system in question stands only a one in ten chance of surviving for as long as $(N_s)_{45}$, i.e. the 90% value. Perhaps more optimistically it stands a 90% chance of lasting for longer than $(N_s)_5$, the 10% value. We can employ a two-tailed approach and say that there is an 80% chance of its lifetime N_s being in the range $(N_s)_5 \leq N_s \leq (N_s)_{45}$.

We could have used the residuals as calculated by Equation (4) instead of the absolute lifetimes for determining percentiles. Indeed, there would have been a great advantage if it could be shown that $(S)_{k\%}$, the residual at the k^{th} percentile, was independent of x, for any k. We would then have been able to use the whole data set as given in Figure 5.5 and compute the k^{th} percentile stability lifetime from

$$(N_s)_{k\%} = (S)_{k\%} (g(x))^{\frac{1}{2}} + g(x) \quad .$$

Alas, this is not the case. The skewness of the distribution increases with x which causes the values of $(S)_{k\%}$ to vary considerably. There is therefore nothing to be gained by normalising the errors in this way.

The percentile approach deals very simply with the problem of the systems which are still observed to be stable although they are in the unstable region. If we are interested in the time within which 90% of all systems are unstable, i.e. $(N_s)_{90\%}$, then all we need do is ensure that any "stable" systems lie in the top 10% before measuring $(N_s)_{90\%}$. If this is not the case then the systems should be numerically integrated until they become unstable or have stability lifetimes exceeding $(N_s)_{90\%}$.

5.5

(ii) The Gamma Distribution

The p.d.f. is an analytical function given by

$$h(t) = \frac{a^b}{\Gamma(b)} \; e^{-at} \; t^{b-1} \qquad (t \geq 0) \qquad (5)$$

where t is equated with time*, a and b are parameters that vary with x, and $\Gamma(b)$ is the Gamma Function

$$\Gamma(b) = \int_0^\infty z^{b-1} \; e^{-z} \; dz \quad , \quad (b > 0) \qquad (6)$$

The cumulative distribution function (c.d.f.) is

$$H(T) = \int_0^T h(t) \; dt \qquad (7)$$

By definition, $H(T) \to 1$ as $T \to \infty$.

The c.d.f. gives the probability of the lifetime of the system being less than T. By substituting Equation (5) into Equation (7) and making the change of variable $u = at$, the c.d.f. may be re-expressed as

$$H(b,U) = \frac{1}{\Gamma(b)} \int_0^U e^{-u} \; u^{b-1} \; du \qquad (8)$$

where $U = aT$.

For a given x, we must find a,b to describe $H(b,U) = H(b,aT)$.

We are then required to solve $H(b,U) = P$ in terms of U for any given probability level P, $(0 \leq P \leq 1)$. For example, if we wish to calculate $T_{90\%}$, the time within which 90% of all systems will become unstable for

*During this discussion we denote the stability lifetime by the simpler T rather than N_s. The units in which the lifetimes are measured are not important at this moment.

5.5

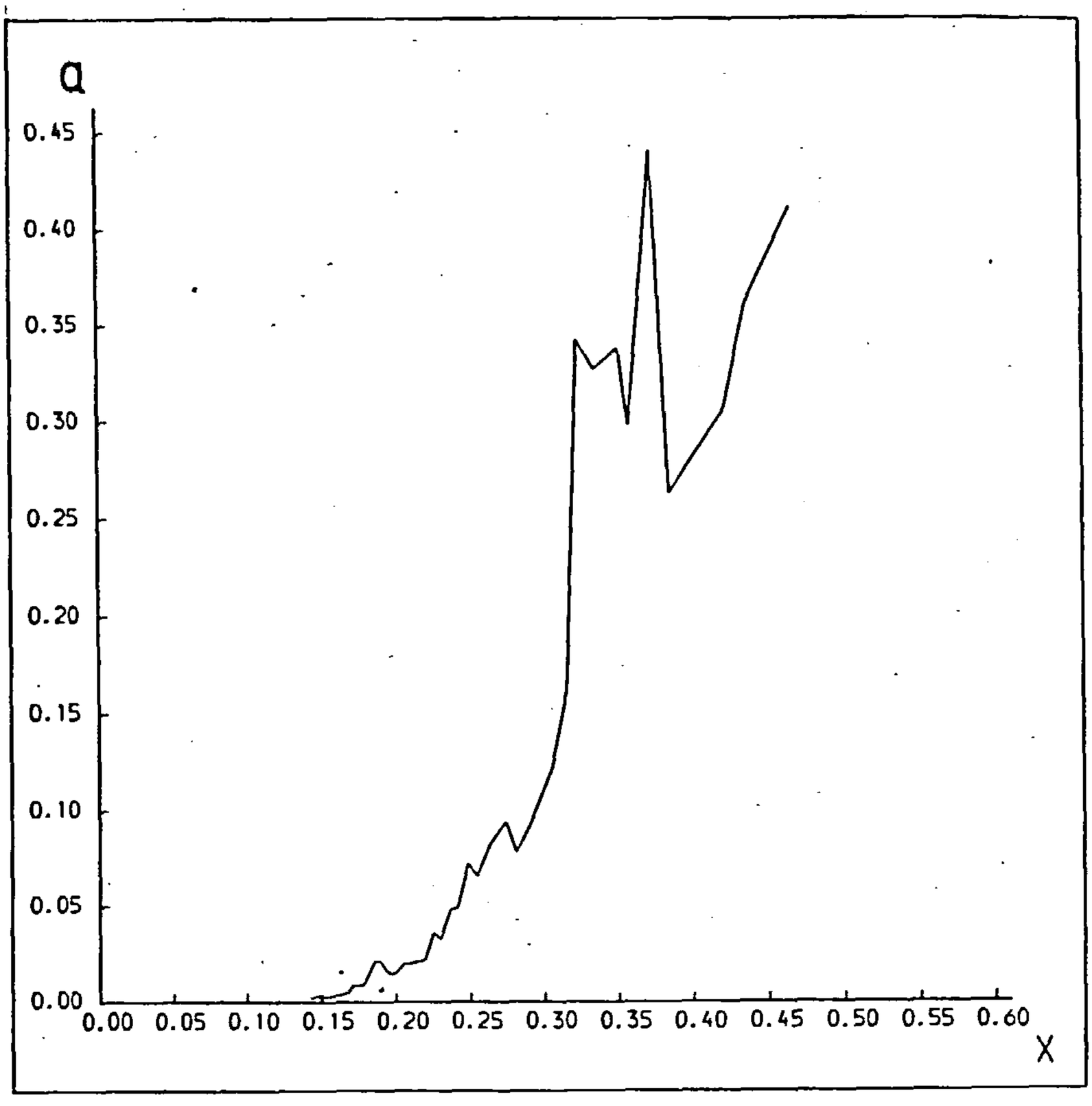
a given x, then we must solve $H(b,U) = 0.9$, in order to obtain U and hence T.

The numerical method for deriving a,b,U for a particular subset of the (x,t) data set is rather complex and is described in Appendix C. We apply the algorithm to small samples taken from the complete (x,t) data set, to derive Figures 5.6 which give a and b against x. The dataset is ordered with increasing x. We take a sample of the first 50 consecutive points, find a,b and assume these values hold at the midpoint in the x-range over which the sample spans. The sample is then moved 10 points (i.e. the first 10 points are excluded, but the 51st-60th points replace them), and the procedure is repeated. This gives sufficient points to plot the graphs in Figures 5.6, but it does mean that the samples are not independent.

On examining the graphs, it is seen that while there is an obvious trend in a, there is scant evidence of any increase or decrease of b with x. The average of b taken from the discrete values is $1.66 \simeq 1.7$. As we shall see below, it is possible to take $b = 1.7$ for all x without affecting the distribution's power to predict stability lifetimes. This assumption of the constancy of b is useful in that it provides an easier algorithm for determining T (Appendix C). Now for a given probability, U need only be calculated once, i.e. it is independent of x. a is also easily calculated since

$$a = b/\bar{t} \qquad (9)$$

where $\bar{t}$ is mean lifetime of the sample. Thus $T = U/a$ can be readily evaluated by a simple calculation of the mean at each sample multiplied by U/b for the particular probability level of interest.



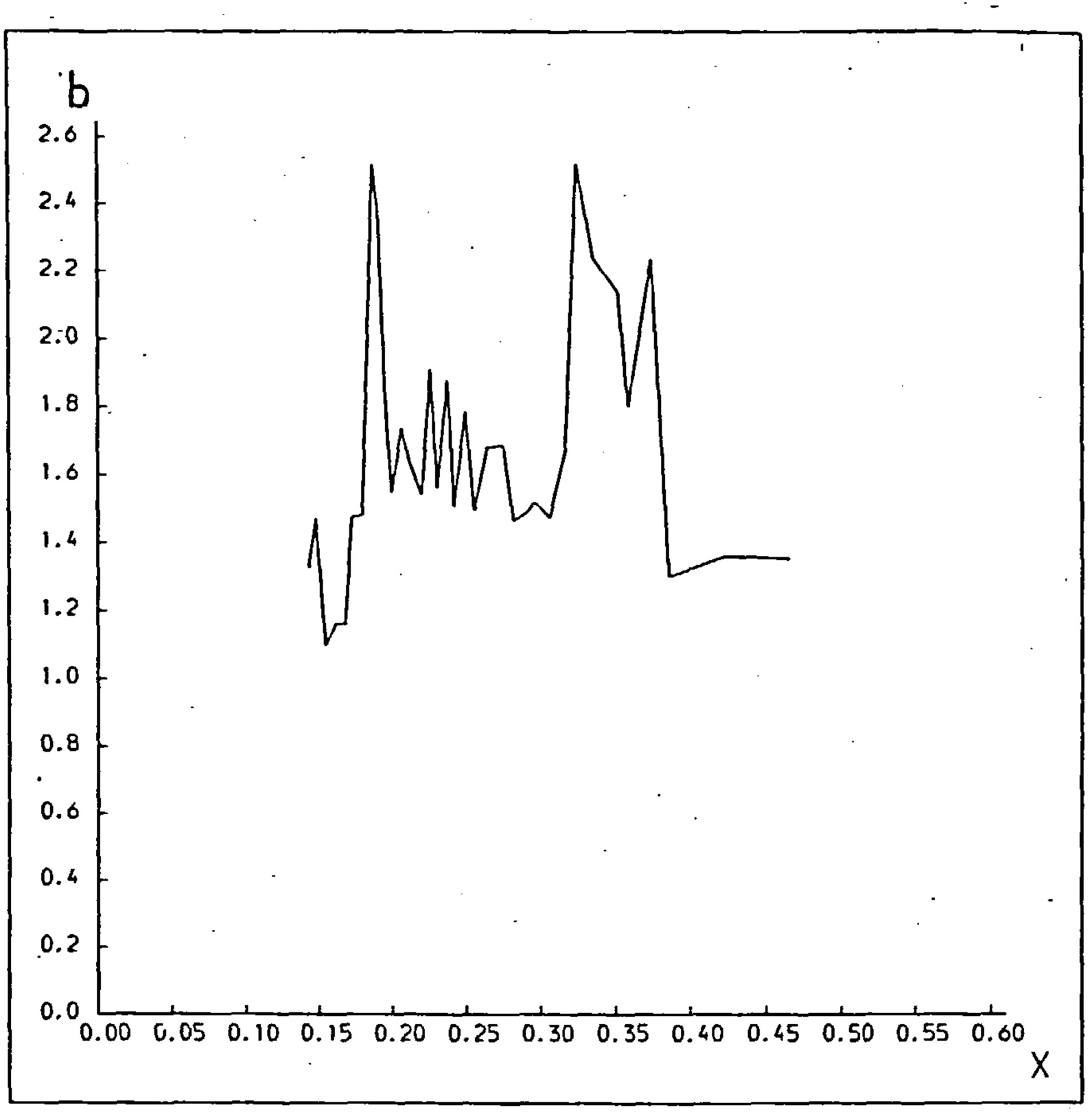


Figure 5.6: Gamma Distribution parameters a, b against x.

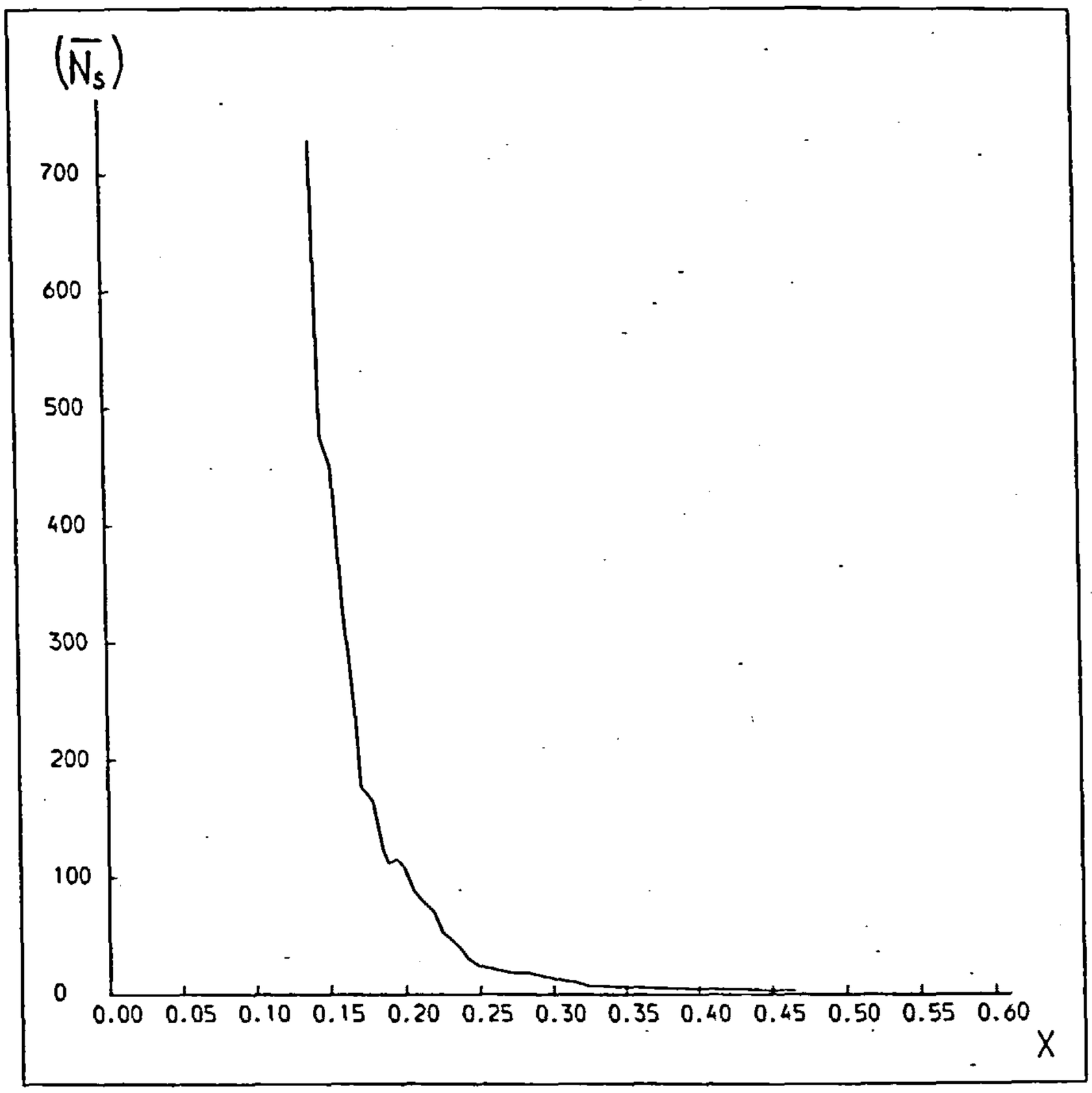


Figure 5.7: Graph of mean stability lifetime $(\overline{N_s})$ against x for b = 1.7.

P	U/b	P	U/b
0	0	0.50	0.812
0.05	0.143	0.55	0.900
0.10	0.225	0.60	0.996
0.15	0.299	0.65	1.102
0.20	0.369	0.70	1.222
0.25	0.438	0.75	1.360
0.30	0.508	0.80	1.526
0.35	0.579	0.85	1.734
0.40	0.653	0.90	2.021
0.45	0.730	0.95	2.499

Table 5.5: U/b at various probability levels P, for b = 1.7.

5.5

When the data is processed with b = 1.7 for all x, the values of U/b are found to be as given in Table 5.5 and the graph of $(\overline{N_s})$ against x is given in Figure 5.7.

There is however one problem still to be overcome. There are the systems which show no signs of instability for long periods although they show $\alpha > \alpha_o$, i.e. there are unstable systems with lower initial α values. We shall call these systems, "Anomalous Stable Systems" or A.S.S.'s, for short. It was seen in Section 5.4 that most ASS's could be associated with some commensurability. In this section we are concerned with their nuisance value when it comes to estimating probability distributions.

A priori, we are unable to determine if any system is an ASS. Indeed we have no formal definition yet of what an ASS is. We cannot say for certain whether an ASS is stable for all time or whether if we run the integration procedure for longer, the system will eventually become unstable. How long must we run a system before we can call it "anomalously stable"? Clearly, the ASS's cannot be used for estimating the p.d.f. as no stability lifetimes are available for them. In what way does their absence affect the distribution?

The definition we finally adopt is the following:

an Anomalously Stable System is defined to be one whose stability lifetime exceeds 3 $T_{90\%}$, where $T_{90\%}$ is the time within which 90% of unstable systems exhibit their instability.

This definition means that some of the systems that have been seen in Figures 5.1 to be unstable are in fact ASS's.

The systems in question are given in Table 5.6.

5.5

ϵ^{23}	ϵ_{32}	α	Stability lifetime (synodic periods)	$3\ T_{90\%}$ (synodic periods)
10^{-3}	10^{-4}	0.805	234.3	119.3
10^{-3}	10^{-5}	0.8225	35.3	24.5
10^{-3}	10^{-6}	0.81231	42.8	28.6
10^{-3}	10^{-6}	0.8	184.6	80.9
10^{-4}	10^{-2}	0.532	236.1	123.4
10^{-4}	10^{-5}	0.92	78.2	40.0
10^{-6}	10^{-3}	0.800192	1481.8	688.5

Table 5.6: Anomalous Stable Systems with measured stability lifetimes from Figures 5.1

For consistent data analysis, the ASS's must be excluded from the data set not only when finding the p.d.f. but also in evaluating $\bar{t}$, in order to calculate U and T. They must also be excluded when fitting the (α, N_s) curves for individual epsilon pairs. Starting from a fresh set of data, we are unable to determine $T_{90\%}$, therefore no unstable systems are excluded. After fitting (α, N_s)-curves, and normalising all data through Equation (2), $T_{90\%}$ may be obtained at any x by the method above. We are then able to see if any of the unstable systems became unstable too late (i.e. $N_s > 3\ T_{90\%}$); if so, they must be excluded. We can also see if there are any stable systems with $N_s < 3\ T_{90\%}$, whose stability is unproven. They must then be studied for longer before they can be classed as an ASS. Having excluded all ASS's from the data set, the whole procedure is repeated. This may have to be repeated several times as the resulting $T_{90\%}$ tends to decrease along with the spread of the available data. However the procedure does converge until a stage is reached where no new ASS's are discovered.

5.5

This full procedure has in fact been carried out for the data presented in this chapter. The curves in Figures 5.1 are fitted excluding the points in Table 5.6. Out of the total dataset of 453 systems with $\alpha > \alpha_o$, 25 are ASS's. There are however 17 systems which are stable after 4000 synodic periods (the numerical limit due to accumulated errors) but 3 $T_{90\%}$ > 4000 synodic periods. These systems cannot be classed as ASS's but are of no use in curve-fitting. Thus the true percentage of ASS's in the dataset is 25/(453-17) = 5.8% ∿ 6%.

Thus in considering the probable lifetime, we must say that there is a 6% chance of the system being an ASS (i.e. exceeding 3 $T_{90\%}$). After that we may go on to assign more quantitative probability levels to other lifetimes.

(iii) The Chi-Square Distribution.

The last choice for a p.d.f. is the Chi-Square Distribution given by

$$h(t) = \frac{1}{2^{n/2}\,\Gamma(n/2)}\, t^{n/2-1} e^{-t/2}, \quad t > 0 \qquad (10)$$

where t is associated with the stability lifetime as before, and n is a single parameter which must be determined from the sample (t_i). On comparing Equation (10) with Equation (5) we can see that the Chi-Square Distribution is a special case of the Gamma Distribution with a = ½ and b = n/2.

On realising this, it is obvious that the Chi-Square Distribution is too restrictive for determining the actual distribution accurately. Figure 5.7 demonstrates that a changes in a marked manner over all possible x and so an assumption that a = ½ is not valid. For this reason, we rule out this particular distribution.

5.5

Comparing the three choices of p.d.f. that have been considered, we see that whereas the Chi-Square Distribution has been ruled out, the Empirical and Gamma Distributions both seem reasonably promising.

In the case of the Gamma Distribution it seems that b may be set as constant over all x.

The lack of an analytic form of p.d.f. in the Empirical Distribution is both an advantage and a disadvantage. The advantage is that with no underlying assumptions made, this method should always allow us to predict stability lifetime intervals regardless of the form of the distribution. The disadvantage is that we need more data to accomplish the same accuracy in prediction. The reason is that by choosing the Gamma Distribution, we are using additional information concerning the form of the distribution. It therefore needs a smaller dataset to find accurate values of a,b from which probable lifetime intervals may be calculated. Without this help from an analytic form, the distribution must be derived only from data.

An analytical form is also more useful when considering the wings of a distribution. For example, given a sample of 50 lifetimes in a neighbourhood of a particular x-value, having evaluated the p.d.f. as a Gamma Distribution, it is possible to estimate $T_{5\%}$ with some accuracy. By the Empirical Method, after ordering the lifetimes in increasing size, $T_{5\%}$ would be estimated as lying between the second and third measured lifetimes. The sparcity of data here will make the estimation error very large, so care must be taken when using the Empirical Method at the extremes of the data.

5.5

One final point on the subject of sample sizes may be made. Although by increasing the sample we can obtain a better estimate of the distribution, if we are still using the same total dataset, the range in x that the sample spans will be greater. The effect will be that "true" distributions from various x-values will be added together and will cumulatively produce a much broader distribution that will not truly reflect the real distribution at any particular x. Therefore the smaller we can make the sample without introducing large errors in the estimation of the p.d.f., the better. This is more easily done with the Gamma Distribution.

The Empirical Method is easier to implement than the Gamma Distribution, but this advantage is lessened considerably if b is allowed to be constant. It is also possible to treat ASS's more easily by the Empirical Method although there may be problems running them for long enough, particular at low x, where an excessively large sample will result in a very false broad distribution as discussed above.

In conclusion, the Gamma Distribution is likely to be more accurate and reliable than the Empirical Method in predicting stability lifetime intervals, particularly at extremes in the data. This is true only if the Gamma Distribution adequately reflects the true distribution of lifetimes for a given x. If this does not turn out to be the case, then we must either find a more suitable analytical form for the p.d.f. or employ the Empirical Method.

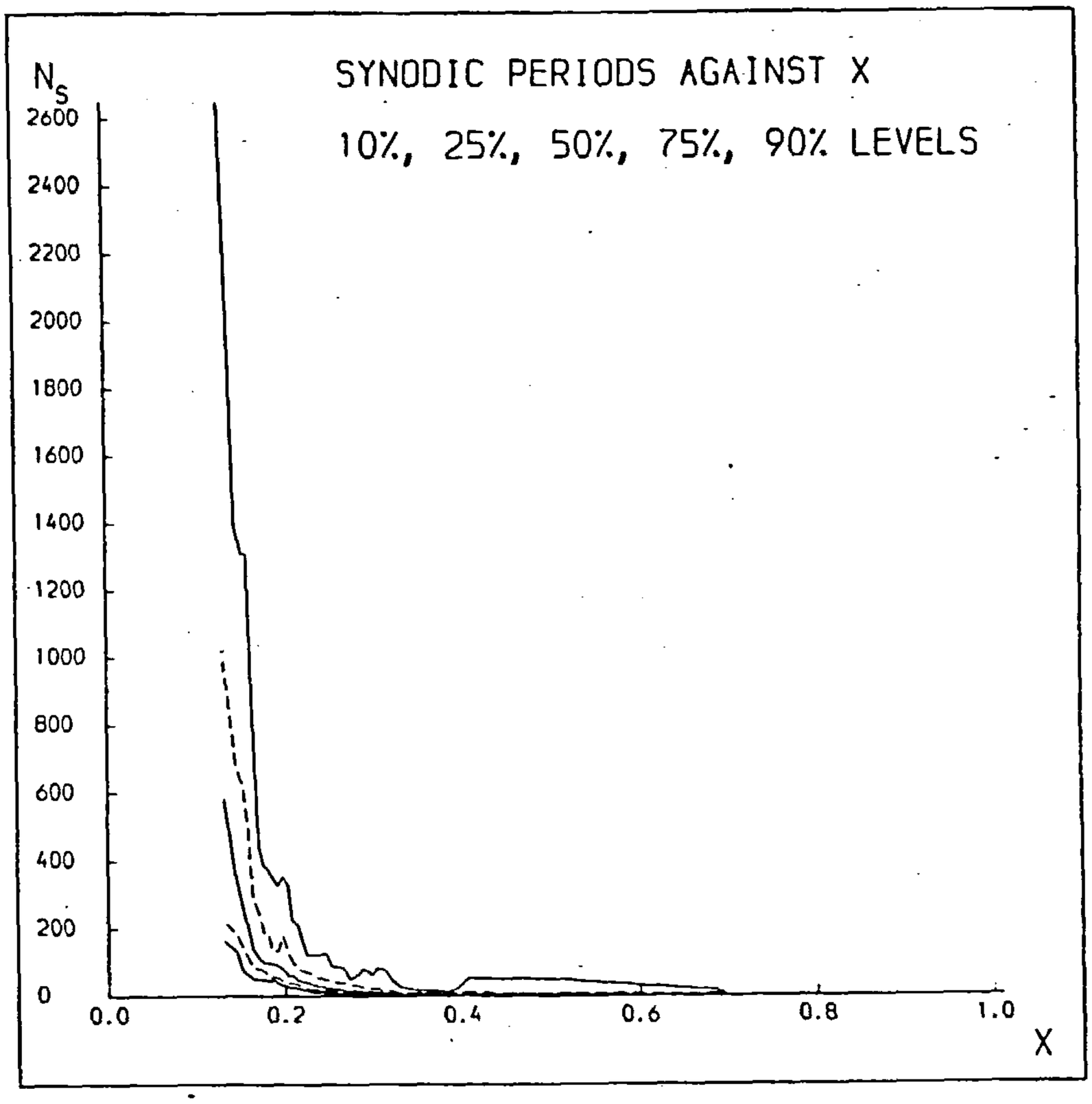
Having discussed at some lengths the merits and demerits of both approaches, we now apply them to the available dataset to derive the actual probability intervals for the stability lifetimes. We shall be concerned with how $(N_s)_{10\%}$, $(N_s)_{25\%}$, $(N_s)_{50\%}$, $(N_s)_{75\%}$ and $(N_s)_{90\%}$ vary

5.5

with x, where $(N_s)_{k\%}$ is the stability lifetime, measured in synodic periods, within which k% of all systems will exhibit instability. Figures 5.8 show the curves $(N_s)_{k\%}(x)$ for the Empirical Method and Gamma Distribution. The Empirical Method is used in Figures 5.8(a) and (b), with sample sizes of 50 and 100 respectively. Figure 5.8(c) is obtained from the Gamma Distribution with both a and b varying as in Figures 5.6. Figure 5.8(d) is obtained from the Gamma Distribution with b = 1.7 and a varying as in Figure 5.7. A sample size of 50 was used for both graphs. In all four cases, the sample was moved with respect to x, by 10 points each time. This means that the points on the curves are not calculated from independent data. Results from the four graphs are presented in tabular form in Tables 5.7.

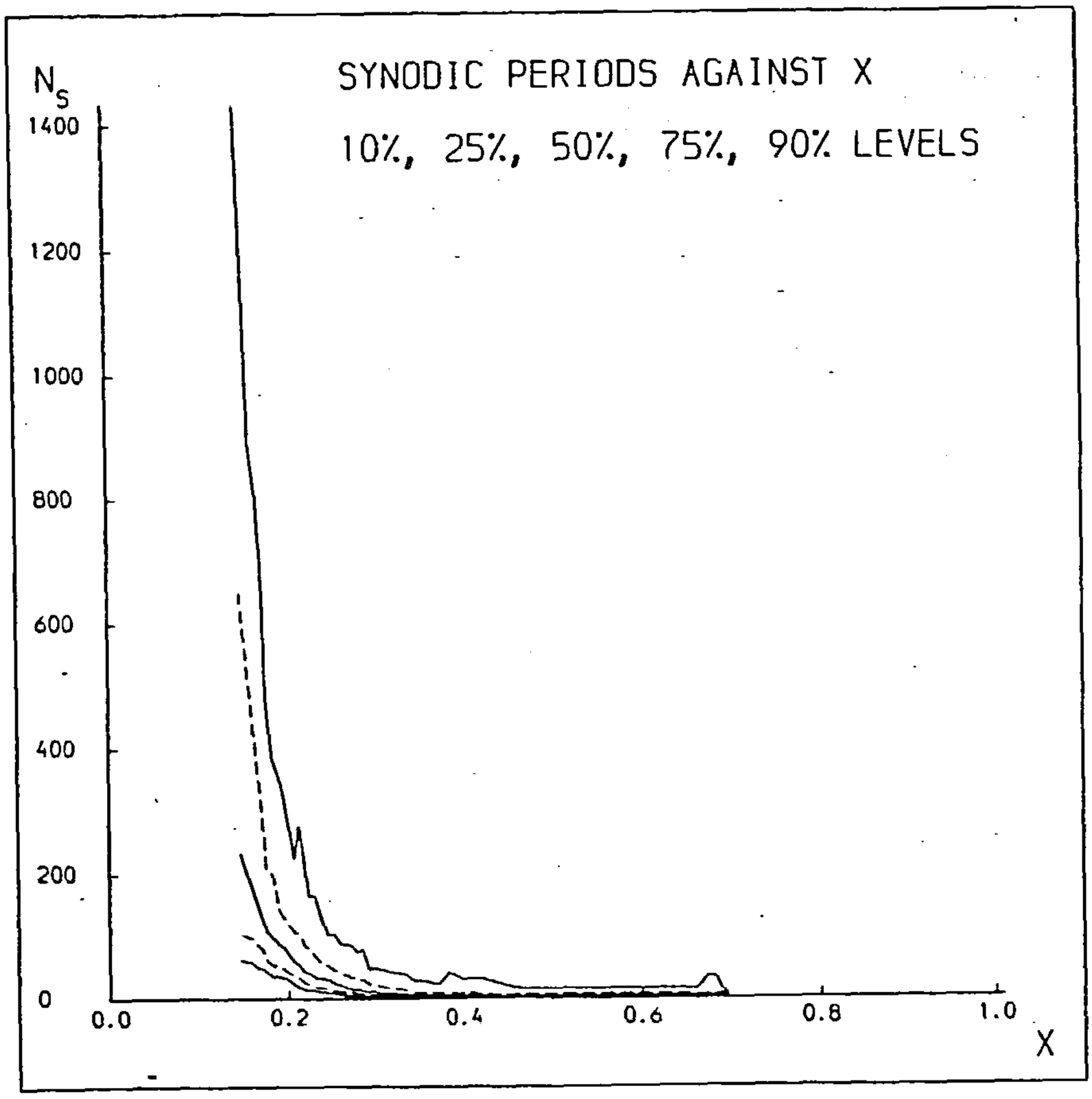
Comparing Tables 5.7 (a) and (b), we observe that the results using the Empirical Method for different sample sizes are comparable. As expected, the larger sample size ensures that the curves in Figure 5.8(b) are smoother than those in Figure 5.8(a). When we look at Tables 5.7(c) and (d), we find that the results are in approximate agreement for the two uses of the Gamma Distribution. The curves in Figure 5.8(d) are smoother than in Figure 5.8(c), due to the fixing of b in the former.

Much greater differences become apparent when we compare the two different methods of curve fitting. The curves for the Empirical Method are noticeably rougher than for the Gamma Distribution. There are two reasons for this. The first is that the Gamma Distribution curves have been processed with the ASS's removed, where as the Empirical Method includes them. This means that the 90% probable lifetime will be particularly rough when ASS's are taken into account. (see Figures 5.8(a) and (b)). The second reason is that the curves for the Empirical

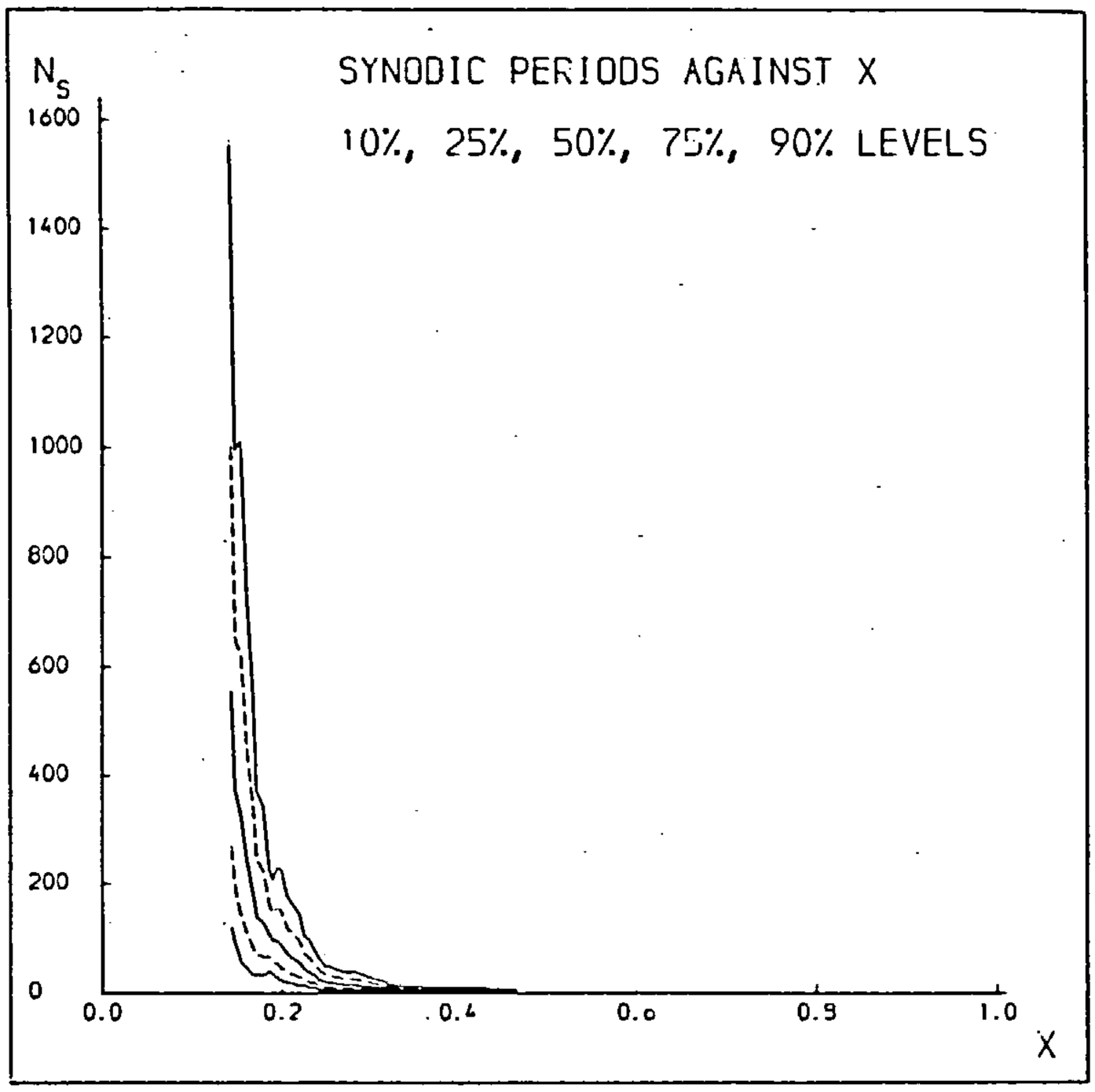


Figures 5.8: Probable Stability Lifetime Curves by:-

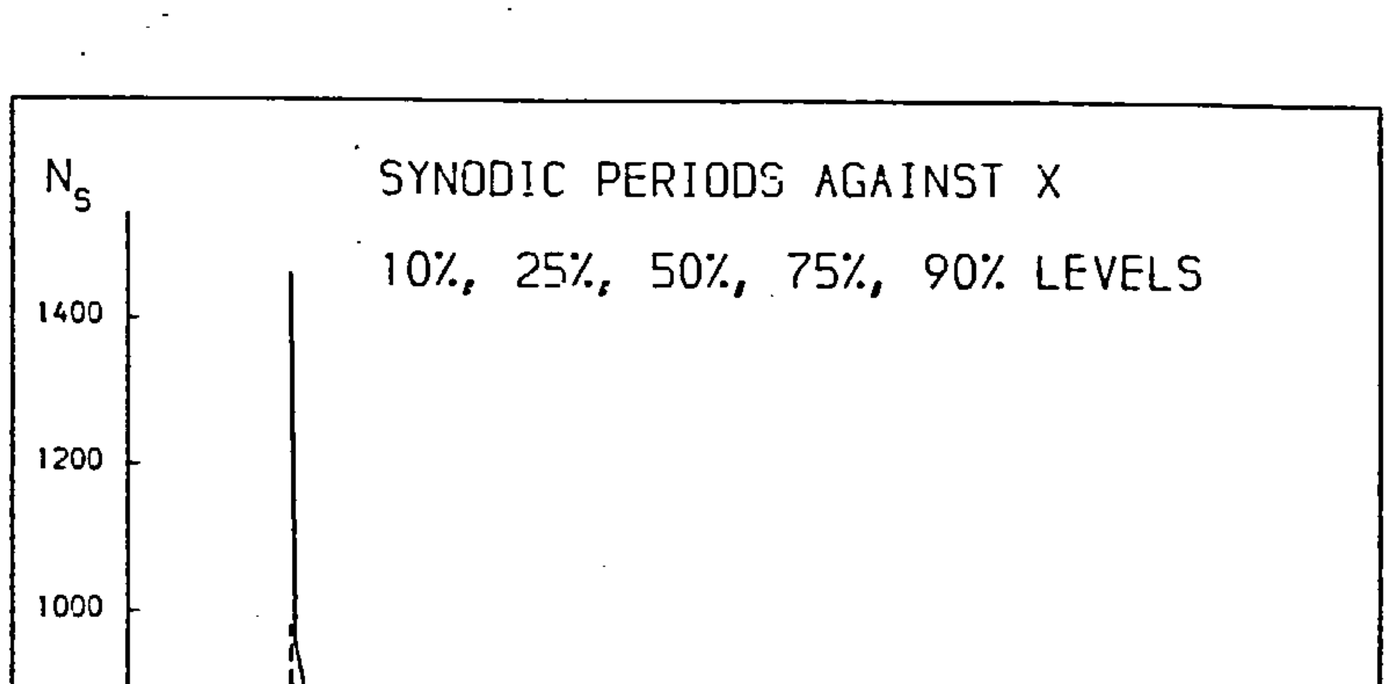
(a) Empirical Method. (Sample size = 50)



(b) Empirical Method. (Sample size = 100)



(c) Gamma Distribution with varying a,b. (Sample size = 50)



(d) Gamma Distribution with varying a, and b fixed as 1.7
(Sample size = 50)

Table 5.7(a) Probable Stability Lifetimes calculated by the Empirical Method with a sample size equal to 50.

x	g(x)	$(N_s)_{10\%}$	$(N_s)_{25\%}$	$(N_s)_{50\%}$	$(N_s)_{75\%}$	$(N_s)_{90\%}$
0.4765	2	0.70	1.11	1.94	5.20	14.7
0.3832	4	1.62	2.04	3.54	7.92	21.7
0.3394	6	1.69	3.42	5.37	8.09	31.3
0.3128	8	3.70	4.55	7.79	19.2	80.2
0.2943	10	4.35	6.81	11.0	19.8	61.9
0.2805	12	4.44	8.98	15.4	34.2	62.3
0.2697	14	4.80	10.1	16.1	35.2	50.8
0.2609	16	4.80	10.3	17.6	35.2	82.8
0.2535	18	4.82	11.2	21.6	39.5	88.2
0.2472	20	8.68	13.5	20.0	40.3	88.2
0.2348	25	11.4	17.2	29.4	54.7	122.
0.2255	30	15.5	21.6	39.6	70.5	122.
0.2182	35	20.6	31.1	42.8	73.4	164.
0.2122	40	21.9	33.0	45.7	82.5	211.
0.2071	45	23.3	36.6	59.4	104.	225.
0.2028	50	26.4	36.8	59.4	112.	305.
0.1990	55	31.5	40.0	66.8	135.	326.
0.1957	60	32.8	48.0	88.2	178.	354.
0.1927	65	36.5	53.0	92.1	131.	326.
0.1900	70	36.5	51.4	92.1	131.	326.
0.1876	75	47.9	57.0	97.4	139.	354.
0.1854	80	47.9	57.0	97.8	139.	326.
0.1833	85	44.5	57.0	97.8	165.	354.
0.1815	90	44.5	58.1	98.7	182.	377.
0.1797	95	47.9	60.2	98.7	182.	377.
0.1781	100	48.1	60.2	98.7	182.	377.
0.1752	110	49.0	74.0	100.	182.	386.
0.1725	120	49.0	74.0	99.4	204.	380.
0.1702	130	49.0	79.9	113.	246.	442.
0.1681	140	49.0	85.1	124.	263.	581.
0.1662	150	49.0	85.1	138.	294.	866.
0.1644	160	55.6	93.2	151.	321.	944.
0.1628	170	62.2	97.3	169.	414.	944.
0.1613	180	62.2	106.	177.	455.	944.
0.1599	190	62.2	114.	190.	455.	944.
0.1586	200	62.2	117.	206.	515.	1310.
0.1574	210	62.2	117.	228.	569.	1399.
0.1563	220	62.2	117.	228.	569.	1399.
0.1552	230	70.2	131.	283.	624.	1399.
0.1542	240	70.2	131.	283.	624.	1399.
0.1532	250	71.4	133.	261.	597.	1310.
0.1523	260	71.4	140.	261.	597.	1310.
0.1515	270	71.4	140.	261.	624.	1310.
0.1506	280	71.4	140.	261.	624.	1310.
0.1498	290	71.4	146.	261.	624.	944.
0.1491	300	77.1	153.	283.	659.	944.

Table 5.7(b): Probable Stability Lifetimes calculated by the Empirical Method with a sample size equal to 100.

x	g(x)	$(N_s)_{10\%}$	$(N_s)_{25\%}$	$(N_s)_{50\%}$	$(N_s)_{75\%}$	$(N_s)_{90\%}$
0.4765	2	0.79	1.63	2.90	5.02	9.79
0.3832	4	1.40	1.92	3.65	5.99	10.1
0.3394	6	1.91	3.42	4.57	7.96	16.7
0.3128	8	3.47	4.47	7.38	12.2	29.8
0.2943	10	3.52	4.55	8.23	15.8	32.7
0.2805	12	4.35	6.08	11.4	21.9	41.0
0.2697	14	4.47	8.05	14.0	25.6	46.9
0.2609	16	4.59	10.5	16.2	31.9	51.1
0.2535	18	5.32	11.2	17.7	33.6	59.6
0.2472	20	7.57	11.7	20.7	38.4	63.0
0.2348	25	11.1	15.9	29.8	51.0	88.3
0.2255	30	13.2	18.6	34.1	68.6	130.
0.2182	35	15.5	25.8	42.1	82.6	165.
0.2122	40	17.3	29.8	48.2	97.4	200.
0.2071	45	19.6	33.7	56.7	102.	200.
0.2028	50	23.3	35.9	58.8	103.	200.
0.1990	55	29.2	41.1	66.8	115.	243.
0.1957	60	32.7	43.5	80.0	123.	243.
0.1927	65	34.8	48.2	81.9	131.	297.
0.1900	70	34.8	48.2	82.4	131.	297.
0.1876	75	35.8	50.7	89.0	144.	297.
0.1854	80	36.9	55.3	97.4	183.	326.
0.1833	85	36.9	56.4	98.7	198.	354.
0.1815	90	43.3	60.9	105.	210.	389.
0.1797	95	43.3	60.9	105.	206.	402.
0.1781	100	43.4	61.8	109.	206.	402.
0.1752	110	44.5	70.3	117.	223.	498.
0.1725	120	48.1	75.9	123.	288.	589.
0.1702	130	50.2	81.7	131.	323.	681.
0.1681	140	50.3	83.3	138.	323.	681.
0.1662	150	53.6	88.0	161.	374.	688.
0.1644	160	56.3	93.8	176.	387.	688.
0.1628	170	57.4	95.8	177.	436.	814.
0.1613	180	58.6	97.4	180.	453.	814.
0.1599	190	58.6	97.4	193.	495.	879.
0.1586	200	58.6	97.4	203.	570.	1310.
0.1574	210	60.8	97.5	208.	584.	1404.
0.1563	220	60.8	97.5	208.	584.	1404.
0.1552	230	61.1	101.	210.	590.	1413.
0.1542	240	61.1	101.	210.	590.	1413.
0.1532	250	61.0	108.	217.	597.	1420.
0.1523	260	64.8	113.	223.	603.	1423.
0.1515	270	70.4	116.	234.	608.	1426.
0.1506	280	70.4	116.	234.	608.	1426.
0.1498	290	70.2	118.	245.	621.	1432.
0.1491	300	70.1	118.	245.	628.	1435.

Table 5.7(c): Probable Stability Lifetimes calculated by the Gamma Distribution with a and b varying. (Sample size = 50).

x	g(x)	$(N_s)_{10\%}$	$(N_s)_{25\%}$	$(N_s)_{50\%}$	$(N_s)_{75\%}$	$(N_s)_{90\%}$	A.S.S.
0.4765	2	0.46	1.07	2.28	4.22	6.68	20.0
0.3832	4	0.89	1.93	3.87	6.86	10.6	31.8
0.3394	6	2.01	3.52	6.00	9.45	13.5	40.5
0.3128	8	2.16	4.58	9.08	16.0	24.4	73.3
0.2943	10	2.98	6.13	11.9	20.6	31.2	93.5
0.2805	12	3.63	7.66	15.1	26.5	40.6	122.
0.2697	14	4.06	8.08	15.3	25.9	38.8	116.
0.2609	16	4.53	8.90	16.6	28.0	41.8	125.
0.2535	18	4.73	9.62	18.5	31.8	48.0	144.
0.2472	20	5.92	11.3	20.5	33.9	49.9	150.
0.2348	25	9.91	18.2	32.4	52.5	76.4	229.
0.2255	30	12.7	23.0	43.4	71.5	105.	315.
0.2182	35	14.8	30.5	58.9	102.	154.	462.
0.2122	40	17.2	34.2	64.4	109.	164.	492.
0.2071	45	21.2	39.7	71.3	117.	171.	513.
0.2028	50	19.4	39.5	75.8	130.	196.	589.
0.1990	55	21.7	44.4	85.7	148.	224.	671.
0.1957	60	25.6	49.9	92.8	156.	232.	695.
0.1927	65	33.7	57.5	95.5	148.	208.	624.
0.1900	70	32.7	55.6	92.5	143.	202.	605.
0.1876	75	36.1	61.7	103.	159.	225.	674.
0.1854	80	39.7	66.0	108.	164.	229.	687.
0.1833	85	39.8	66.4	108.	165.	230.	691.
0.1815	90	30.3	64.0	127.	222.	340.	1021.
0.1797	95	30.8	64.6	127.	222.	339.	1018.
0.1781	100	32.4	67.3	132.	229.	348.	1045.
0.1752	110	34.0	69.3	133.	229.	346.	1039.
0.1725	120	34.3	69.5	133.	228.	344.	1033.
0.1702	130	34.6	73.2	145.	254.	390.	1169.
0.1681	140	32.7	78.6	172.	323.	515.	1546.
0.1662	150	32.7	84.9	197.	384.	629.	1886.
0.1644	160	36.1	92.5	213.	413.	673.	2019.
0.1628	170	43.2	106.	237.	450.	724.	2172.
0.1613	180	43.7	108.	240.	457.	736.	2207.
0.1599	190	48.0	115.	251.	470.	750.	2249.
0.1586	200	48.2	120.	268.	510.	823.	2470.
0.1574	210	46.7	126.	299.	594.	982.	2945.
0.1563	220	46.7	126.	299.	594.	982.	2945.
0.1552	230	55.8	142.	326.	631.	1027.	3080.
0.1542	240	55.8	142.	326.	631.	1027.	3080.
0.1532	250	61.1	150.	334.	635.	1022.	3065.
0.1523	260	62.6	152.	337.	637.	1022.	3065.
0.1515	270	61.4	150.	333.	630.	1013.	3039.
0.1506	280	61.4	150.	333.	630.	1013.	3039.
0.1498	290	67.8	158.	337.	623.	985.	2955.
0.1491	300	75.6	170.	351.	636.	994.	2983.

Table 5.7(d): Probable Stability Lifetimes calculated by the Gamma Distribution with varying a and b fixed at 1.7. (Sample size = 50).

x	g(x)	$(N_s)_{10\%}$	$(N_s)_{25\%}$	$(N_s)_{50\%}$	$(N_s)_{75\%}$	$(N_s)_{90\%}$	A.S.S.
0.4765	2	0.69	1.34	2.48	4.16	6.18	18.5
0.3832	4	1.13	2.19	4.06	6.80	10.1	30.3
0.3394	6	1.59	3.09	5.73	9.60	14.3	42.8
0.3128	8	2.62	5.09	9.44	15.8	23.5	70.5
0.2943	10	3.38	6.58	12.2	20.4	30.4	91.1
0.2805	12	4.36	8.48	15.7	26.3	39.1	117.
0.2697	14	4.28	8.32	15.4	25.8	38.4	115.
0.2609	16	4.63	9.01	16.7	28.0	41.6	125.
0.2535	18	5.24	10.2	18.9	31.6	47.0	141.
0.2472	20	5.63	11.0	20.3	34.0	50.5	151.
0.2348	25	8.76	17.0	31.6	52.9	78.6	236.
0.2255	30	11.9	23.1	42.8	71.7	107.	320.
0.2182	35	16.7	32.5	60.3	101.	150.	450.
0.2122	40	18.1	35.1	65.1	109.	162.	486.
0.2071	45	19.4	37.8	70.1	117.	174.	523.
0.2028	50	21.4	41.7	77.3	129.	192.	577.
0.1990	55	24.3	47.3	87.7	147.	218.	655.
0.1957	60	25.8	50.1	92.9	156.	231.	694.
0.1927	65	25.0	48.6	90.0	151.	224.	672.
0.1900	70	24.2	47.1	87.3	146.	217.	652.
0.1876	75	26.9	52.4	97.1	163.	242.	725.
0.1854	80	27.9	54.2	101.	168.	250.	750.
0.1833	85	28.0	54.5	101.	169.	252.	755.
0.1815	90	36.5	71.0	132.	220.	328.	983.
0.1797	95	36.5	71.0	132.	220.	328.	983.
0.1781	100	37.6	73.2	136.	227.	338.	1013.
0.1752	110	37.8	73.4	136.	228.	339.	1016.
0.1725	120	37.6	73.2	136.	227.	338.	1013.
0.1702	130	41.8	81.3	151.	252.	375.	1125.
0.1681	140	52.6	102.	190.	317.	472.	1415.
0.1662	150	62.5	122.	225.	377.	561.	1682.
0.1644	160	67.1	131.	242.	405.	602.	1807.
0.1628	170	73.2	142.	264.	442.	657.	1971.
0.1613	180	74.4	145.	268.	449.	667.	2002.
0.1599	190	76.6	149.	276.	462.	687.	2061.
0.1586	200	83.1	162.	299.	502.	745.	2236.
0.1574	210	96.5	188.	348.	583.	866.	2598.
0.1563	220	96.5	188.	348.	583.	866.	2598.
0.1552	230	103.	200.	370.	619.	921.	2762.
0.1542	240	103.	200.	370.	619.	921.	2762.
0.1532	250	103.	201.	373.	624.	928.	2783.
0.1523	260	104.	202.	374.	626.	931.	2792.
0.1515	270	103.	200.	370.	620.	921.	2763.
0.1506	280	103.	200.	370.	620.	921.	2763.
0.1498	290	102.	198.	366.	614.	912.	2736.
0.1491	300	104.	202.	375.	628.	934.	2801.

5.5

Method reflect the raw data more closely. By fitting an analytic function in the case of the Gamma Distribution Method, the data has undergone additional smoothing. The $(N_s)_{10\%}$ and $(N_s)_{90\%}$ curves for the Empirical Method are particularly rough. This is a reflection of the point made earlier concerning this method's ability to deal accurately with data at either end of the distribution. We would be unwise to use any curves below the 10% limit or above the 90% limit without increasing the sample size. Unfortunately we then run into the problem of "distribution smudging" described earlier.

The tables may be used to predict probable stability lifetimes. For a given system prescribed by ε^{23} , ε_{32} , α, a most likely stability lifetime is given by g(x), where

$$g(x) = f(\alpha;\alpha_o,\beta,\gamma) = \exp\left[\beta\ (\frac{1-\alpha}{\alpha-\alpha_o})^{\gamma}\ \right]-1, \quad (\alpha_o < \alpha \leq 1)$$

as before, with α_o,β,γ as given in Table 5.1 from Section 5.3. We may then read off any prescribed confidence interval from one of the tables, depending on the choice of method. Interpolation may be possible between the values given, but extrapolation should not be attempted as the limits of the tables are also the limits of the dataset where the errors in predictions will be greatest.

For the Empirical Method, the median curve $(N_s)_{50\%}(x)$ is comparable with g(x) as expected, while the median curve for the Gamma Distribution is higher. There does seem to be a difference between corresponding stability curves for the two methods. It may be hoped that with a larger dataset, these differences might be lessened.

In order to test the reliability of these predictions, 100 new direct systems were numerically integrated. These independent test

5.5

systems have ε^{23} , ε_{32} , α spanning the complete ranges already studied, i.e. $10^{-6} \lesssim (\varepsilon^{23}, \varepsilon_{32}) \lesssim 10^{-2}$, $\alpha_o < \alpha < 1$. Care was taken not to study systems that were close to the ends of the available dataset. Each system was integrated until it exhibited instability or until it could be designated as an ASS. It is possible to compare the actual measured stability lifetime with the probable lifetimes given in Tables 5.7. In this way we may assign each system to one of the intervals $(N_s)_{0\%} - (N_s)_{10\%}$, $(N_s)_{10\%} - (N_s)_{25\%}$, etc. We are able to compare the actual numbers of systems lying in the intervals with the numbers expected. For example, out of the 100 systems we would expect half to lie in the range $(N_s)_{25\%} - (N_s)_{75\%}$ barring the presence of ASS's. As was said above, ASS's are more easily accommodated in the Empirical Method and it should be possible to combine them within the $(N_s)_{90\%} - (N_s)_{100\%}$ interval, i.e. all systems with lifetimes exceeding $(N_s)_{90\%}$. For the Gamma Distribution, the ASS's must be treated as a separate class, after which the probable lifetime intervals may be constructed. Thus the expected number of systems with lifetimes in the interval $(N_s)_{25\%} - (N_s)_{75\%}$ is not 50 out of 100 but 47 out of 94 (6% should be ASS's).

The results of these experiments are given in Table 5.8 for the Empirical Method and Table 5.9 for the Gamma Distribution. It is possible to test the goodness of fit of the experiments with the theoretical predictions, by using the Chi-Square Test. The statistic

$$\chi^2_{n-1} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

may be calculated, where n is the number of intervals,

Interval	0-10%	10-25%	25-50%	50-75%	75-90%	90-100%	χ^2_5
Expected	10	15	25	25	15	10	
Actual (Sample Size=50)	13	17	12	27	23	8	12.75
Actual (Sample Size=100)	12	20	14	25	20	9	8.67

χ^2_5 = 11.1 at the 95% level.

Table 5.8: Frequency with which 100 test systems fall into probable lifetime intervals using the Empirical Method taking samples of 50 and 100 from the original dataset.

Interval	0-10%	10-25%	25-50%	50-75%	75-90%	90-3(90%)	ASS	χ^2_6
Expected	9.4	14.1	23.5	23.5	14.1	9.4	6.0	
Actual (varying a,b)	12	20	16	22	12	11	7	6.43
Actual (constant b=1.7)	9	18	22	20	12	12	7	2.91

χ^2_6 = 12.6 at the 95% level.

Table 5.9: Frequency with which 100 test systems fall into probable lifetime intervals using the Gamma Distribution. The lifetime sample size is 50.

O_i is the observed number and E_i the expected number of systems with lifetimes in the i^{th} interval. We are testing the null hypothesis that the models are true predictors of the actual behaviour of direct three-body systems. With the given uncertainties, we should clearly not expect the observed counts to match the expected counts in every case, (giving $\chi^2_{n-1} = 0$, for all samples). The question is how great a discrepancy can we allow between theory and observation? Assuming the null hypothesis, the χ^2_{n-1} statistic has a sampling distribution given by the Chi-Square p.d.f. of Equation (10). The cumulative distribution function (c.d.f) of this distribution is tabulated for various n in Appendix A. The solution of the c.d.f. at some probability level gives the value of χ^2_{n-1} which must be compared with the observed value.

We have chosen the 95% point of the distribution. For the Empirical Method, there are six intervals, therefore 5 degrees of freedom. The solution of the c.d.f. when equal to 0.95 is 11.1. This means that if the null hypothesis is true, there is a 95% probability that the observed χ^2_{n-1} statistic will be less than this value. If it actually exceeds this value, it is deemed to be sufficiently unlikely to cast doubts on the underlying hypothesis. We would therefore consider that the fit was not good enough for us to be able to assume that our model was accurate.

Turning our attention to the tables, we see that the data fits the Gamma Distribution model exceedingly well, both with varying and with constant b. Indeed the agreement between theory and observation is marginally better for the distribution with constant b. There seems little point in using a more complicated distribution than is necessary so the constant b distribution is preferred.

5.5

On the other hand, the Empirical Method does not fare as well, particularly when using a sample size of 50 to estimate the various percentiles. In this case the Chi-Square test fails the method at the 95% level. When we increase the sample size to 100 the agreement is better and the method passes the test, but the results are not as good as for the Gamma Distribution. The small sample of 50 causes severe inaccuracies in prediction. There is an improvement by increasing the sample size but at the cost of distribution smudging over a wider range in x. The remedy of course is to increase the original dataset and herein lies the Empirical Method's greatest drawback; namely the need to acquire considerably more data than the Gamma Distribution to achieve the same accuracy in prediction.

It should be noted that the actual distribution is not exactly modelled by a Gamma Distribution. If it was, then we should expect around 1% of all systems to have stability lifetimes exceeding $3(N_s)_{90\%}$, (i.e. to be ASS's). Instead the figure is nearly 6%, indicating a much larger tail to the distribution. We therefore need the definition of an ASS in this form to eliminate the anomalous tail, in order that we might better fit a Gamma Distribution to the rest of the data.

5.6 Summary

In this chapter, coplanar direct three-body systems have been studied in the same manner as for the retrograde systems. From numerical integration studies, graphs of stability lifetime against initial α were displayed for various combinations of ε^{23}, ε_{32}, in Section 5.2. In general the lifetimes increase as α is decreased from 1, and seem to rise asymptotically to some critical value α_o. The rise is not smooth as the points are scattered widely making any accurate curve-fit difficult.

In Section 5.3, curves were fitted to the data provided by the unstable systems. Unlike the retrograde systems, no use can be made of systems that exhibit no instability in this fitting procedure, but all the data from the unstable systems must be used. The form of the curve that was fitted, given by Equation (1), is parameterised by α_o, β, γ for a given ε^{23}, ε_{32}. For given ε^{23}, ε_{32}, α_o; β and γ may be found by calculating u and v for each data point (α, N_s), (Equations (4.4) and (4.5))and constructing the least squares fit. The (u,v) points are weighted with increasing u. The goodness of fit of the curve for α_o, β, γ may be found by computing the correlation coefficient r (Equation (4.6)). The value of α_o (with corresponding β and γ) is varied in the range (α_c, α_{us}). The best fitting values of α_o, β, γ maximise r.

Section 5.4 discussed the role commensurabilities play in causing the scattering of the (α, N_s) points about the best fit curve. It was seen that in most cases of anomalously stable systems, there was a nearby commensurability that could account for the behaviour. It was also seen that there was a significant preponderance of "odd-spoked" commensurabilities compared to "even-spoked" commensurabilities which helped stabilize systems in this way. Occasionally there was evidence of a much broader range of anomalously stable systems in the same manner as

5.6

the retrograde systems (discussed in Chapter 4).

The deviations from the best fit curve make accurate predictions very difficult. Section 5.5 was concerned with methods of obtaining predictions of quantifiable uncertainty. To this end, the (α,N_s) data from each ε^{23} , ε_{32} pair may be combined into one data set (x,N_s) by normalising α according to Equation (2).

For any system of this kind that we may wish to study, we may calculate from ε^{23}, ε_{32}, α , a value of x*. By considering only that part of the (x,N_s) data set in the immediate neighbourhood of x*, we may fit a probability distribution to the (N_s) data that allows us to predict possible lifetime ranges. For example we can compute $(N_s)_{90\%}$ which gives the time within which 90% of all systems characterised by value x* should become unstable.

Three models for the probable lifetime distribution were considered. These used the Empirical Method, Gamma Distribution and Chi-Square Distribution as described in Appendix C. The Chi-Square Distribution, being a one parameter function was too restrictive. Of the remaining two, the Gamma Distribution is preferred. Although the Empirical Method always is more widely applicable and easier to use, for this particular dataset the Gamma Distribution has proved more reliable and accurate in its subsequent predictions of test systems. It is considerably simplified since one of its parameters, b, may be assumed to be constant at 1.7 without affecting the accuracy.

From the probability distributions, we were able to class certain systems as "anomalously stable" if their stability lifetimes exceeded 3 times the 90% lifetime limit. By neglecting these systems, the curve-fitting of α_o,β,γ can be repeated with a smoothed data set which in

turn allows more accurate normalisation to x. With more accurate normalisation the predictions are liable to be constrained over a smaller range.

This definition of the "anomalous stable system" has already helped in compiling the data set as it highlighted stable systems which should be studied for longer. Many of these systems subsequently became unstable within $3(N_s)_{90\%}$. Thus some of Walker and Roy's original data has been modified and the empirical stability region has shrunk slightly from their original estimates. It would therefore be prudent to perform many more experiments around $\alpha_o(\varepsilon^{23}, \varepsilon_{32})$ to determine its value accurately.

A final point follows from this. We may expect that by increasing the available data set, we will vary our estimates of α_o, β, γ , hopefully in such a way as to achieve a better fitting curve to the data. With more accurate α_o, β, γ we should expect that normalisation of α to x will result in the spread of points about the curve

$$g(x) = \exp\left(\frac{1}{x} - 1\right) - 1$$

being less, in other words, the probable lifetime ranges should be smaller in width. (Note that the reliability of the predictions should not improve; only the precision of the predictions that we make). The spread of (x, N_s) points will shrink with increasing data to some limit, where the spread is due only to physical considerations such as commensurabilities, and not to statistical errors in the distribution fitting.

Chapters 4 and 5 have been concerned with numerical studies of fictitious three-body systems. In the next chapter, we shall compare the results for fictitious systems with real three-body subsystems contained in the Solar System.

CHAPTER 6

COMPARISON OF DIRECT AND RETROGRADE THREE-BODY SYSTEMS

6.1 Introduction

The previous three chapters have been concerned with aspects of the hierarchical stability of general three-body systems. Chapter 3 was concerned with the sufficient conditions to preserve the hierarchical arrangement of the three bodies. Chapter 4 dealt with fictitious retrograde systems, while Chapter 5 dealt with fictitious direct systems. In both these chapters, the emphasis was on the processing of data derived from numerical integration experiments. The systems studied were in no way related to real subsystems of three bodies present in the Solar System. It was seen that for both direct and retrograde cases, there was a considerable region in the configuration space where systems were hierarchically stable. This empirical stability region was seen to enclose the analytical region of hierarchical preservation, described in Chapter 3.

By way of a summary of the preceding chapters, we shall be concerned with an overall comparison of these results, as well as comparing them with various three-body subsystems of the Solar System.

In Section 6.2, we shall compare the empirical stability regions for direct and retrograde systems as well as comment on the differences between them and the stability regions predicted by the analytical "Hill-type" stability involved in preserving the hierarchy. Section 6.3 will comment briefly on the effects of commensurabilities on direct and retrograde systems.

Walker and Roy, in Paper I of their series, published extensive tables of the values of ε^{23} and ε_{32} for three-body subsystems in the Solar System as well as for triple star systems and fictitious systems investigated by other authors. These tables were drawn up in 1980.

6.1

Since then, our knowledge of the physical and mineralogical properties of the planets and satellites has increased dramatically. Most notably, the Voyager programme has discovered many new satellites in the Jovian and Saturnian systems as well as returning more accurate data concerning those satellites that had been discovered from terrestrial observations. For the first time the physical dimensions and mineral composition of the small satellites can be accurately determined, leading to measurements of the masses which are considerably better than the guesses of only a decade ago. More knowledge has also been gained concerning the orbital characteristics of some of these bodies.

Land based observations have also improved. In 1978, it was noted that photographs of Pluto showed a slight elongation. Using the techniques of speckle interferometry, Bonneau and Foy in 1980 confirmed the existence of Charon, a satellite of Pluto. Spectroscopic observations of Pluto have also shown it to be covered in frozen methane rather than rock. Its albedo is therefore much greater than was previously thought. Given its apparent magnitude as seen from the Earth, we now find that Pluto is between 2000 and 3000 km in diameter (less than the Moon). It therefore seems that its mass is very much smaller than first estimated.

Given that our knowledge of the Solar System has increased so dramatically, we repeat the work of Walker and Roy in Section 6.4 and determine ε^{23}, ε_{32}, α_c, α_o for various subsystems, using revised estimates for the masses of the planets and satellites. (This is done in the certain knowledge that the data on Uranus are immediately out of date due to the fly-past by Voyager). Comments are made concerning the stability of the relevant systems, particularly where retrograde

6.1

satellites are involved.

In Section 6.5, we discuss the implications that this work has for possible theories concerning the dynamical origin of the Solar System.

6.2 A Review of the Numerical Experiments

In this section, we shall compare the behaviour of the direct and retrograde three-body systems numerically integrated in Chapters 4 and 5. In particular, the results from Figures 4.1 and Table 4.1 (for retrograde systems) will be compared with Figures 5.1 and Table 5.1 (for direct systems). Some comparisons have already been made within the preceding chapters; where necessary, these will be summarised while presenting a suitable overview of all the experiments.

All the three-body systems studied have been coplanar, initially circular and starting from a conjunction of m_1 - m_2 - m_3. These systems can be classified unambiguously by their values of ε^{23} , ε_{32} , α and the rotational sense of the outer binary (M_2,m_3) with respect to the inner binary (m_1,m_2), i.e. direct or retrograde motion.

In Chapter 3, it was shown that, given ε^{23}, ε_{32} , there existed α_c such that all systems defined by ε^{23} , ε_{32} , $\alpha < \alpha_c$ would retain their initial hierarchy for all time. The surface $\alpha_c = \alpha_c (\varepsilon^{23}\ \varepsilon_{32})$ differed depending on whether the systems in question were direct or retrograde. It was seen that for direct systems, $\alpha_c \to 1$ as ε^{23} , $\varepsilon_{32} \to 0$. Over the range of ε^{23},ε_{32} studied, α_c rose monotonically with ε^{23} , (ε_{32} fixed) and with ε_{32} (ε^{23} fixed). This was entirely in line with the intuitive nature of the problem, namely, that as the perturbations decrease, m_2 should be able to approach closer to m_3 without endangering the stability of the system.

6.2

In retrograde systems however, we found that the corresponding α_c surface did not reflect this intuitive idea. Instead, the highest value of α_c occurred when $\varepsilon^{23} \sim \varepsilon_{32}$. For many pairs of $(\varepsilon^{23}, \varepsilon_{32})$, there were no values of α_c, and for every pair, the value of α_c for the retrograde case was considerably less than α_c for the direct case.

Considerable use has been made of this analytical criterion for hierarchical preservation in the case of direct systems. Its use for retrograde systems was far less clear. Either the α_c surface does not accurately reflect the physical behaviour of the system, or our intuitive ideas concerning the empirical stability parameters are wrong.

This question was resolved in Chapters 4 and 5. From many hundreds of numerical integrations carried out on fictitious direct and retrograde systems, graphs of stability lifetime N_s against α were obtained for given $(\varepsilon^{23}, \varepsilon_{32})$ pairs. For each pair, as α decreased from 1, N_s was seen to rise, tending to infinity at some value of α equal to α_o. Wherever possible a curve

$$N_s = f(\alpha) = \exp\left[\beta \left(\frac{1-\alpha}{\alpha-\alpha_o} \right)^{\gamma} \right] - 1 \qquad (1)$$

was fitted in the range $(\alpha_o, 1\ ]$. By a statistical approach, values of α_o, β and γ were found which gave the best fit to the available data. While some physical meaning can be attached to α_o, there is no physical interpretation of either β or γ . In fact, there is scant evidence of any systematic trends in β or γ as functions of ε^{23} and ε_{32} , as can be seen from Tables 4.1 and 5.1.

There are, however, systematic trends in $\alpha_o = \alpha_o\ (\varepsilon^{23}, \varepsilon_{32})$ that apply to both direct and retrograde systems. If we fix ε^{23} and decrease ε_{32}, α_o rises to some limit. This limit itself varies with

the value that ϵ^{23} is fixed at, but it tends to unity as ϵ^{23} tends to zero. If we study Table 4.1, for retrograde systems, we see that the limit is attained when $\epsilon_{32} \lesssim \epsilon^{23}$.

A similar situation occurs if we fix ϵ_{32} and decrease ϵ^{23} . α_o rises to a limit as before and this limit tends to unity as ϵ_{32} tends to zero. From Table 4.1, this limit is attained when $\epsilon^{23} < \epsilon_{32}$, but is not as well maintained. Indeed there is a little evidence to suggest that a peak in α_o is attained when $\epsilon^{23} = \epsilon_{32}$ and thereafter the value α_o drops slightly as ϵ^{23} is decreased. It may be that this effect is due only to statistical fluctuations in fitting α_o.

In the case of direct systems, similar trends probably exist, but it is difficult to be certain about this because the commensurabilities allow for considerable errors in determining α_o, as was discussed in Section 5.4.

It certainly seems, both from this comparison and also from the individual graphs of Figures 4.1, that the systems which show $\epsilon^{23} = \epsilon_{32}$ are critical in their behaviour, especially for the retrograde systems. They may be divided into classes: those with $\epsilon^{23} < \epsilon_{32}$ and those with $\epsilon^{23} \gtrsim \epsilon_{32}$. With the latter class, it is pointless fitting curves in the domain $(\alpha_o, 1]$ since for a given pair of ϵ^{23} , ϵ_{32} the domain D of α over which N_s rises from one to infinity is vanishingly small. When $\epsilon^{23} < \epsilon_{32}$, the rise is more leisurely and shows the step nature (Figures 4.1), allowing curves to be fitted through the centre of each step.

It is interesting to note that one of the reasons why Walker and Roy chose to group systems by ϵ^{23} , ϵ_{32} rather than the normalised masses μ, μ_3 , was to produce a sharper cut-off between stable (infinite

6.2

N_s) and immediately unstable systems ($N_s < 1$) for a graph of N_s against α . For direct systems with given ε^{23}, ε_{32}, the domain D is usually large enough to attempt a curve fit with some success as the cut-off is never that sharp. However for retrograde systems with $\varepsilon^{23} \geq \varepsilon_{32}$, the full potential of the stability parameters in this respect is realised.

For retrograde systems with $\varepsilon^{23} < \varepsilon_{32}$, it is possible to fit curves of the type given in Equation (1) to the (α, N_s) data. The data is very much smoother for the retrograde systems than it is for the direct. (It should be recalled from Chapter 4 that the general behaviours of unstable retrograde systems were remarkably similar compared to the direct systems). There were considerable uncertainties when curves were fitted to the direct systems. Because of the smoothness of the retrograde data the uncertainties in curve-fitting were much reduced. It became apparent however that Equation (1) was not an entirely suitable form for fitting to the data. A much more complicated function is probably needed to get more accurate results (see Chapter 8). For the time being, we must be cautious about our findings concerning the region of empirical stability. The systematic errors involved in evaluating α_o from Equation (1) prompt us to express our findings in the following way.

- *Insofar as $f(\alpha)$, as given in Equation (1), models the stability lifetimes of systems in the domain $(\alpha_o, 1]$, for given ε^{23} , ε_{32} :-*

(i) there exists a value α_o, such that any system with $\alpha < \alpha_o$ is hierarchically stable.

(ii) α_o always exceeds α_c, the critical value of α below which a system has an analytical guarantee of hierarchical preservation.

(iii) α_o *for retrograde systems exceeds* α_o *for direct systems,* $\forall \varepsilon^{23}, \varepsilon_{32}$.

These are the major findings from Chapters 4 and 5, so it is worth looking at them in more detail. From the first point, within the $(\varepsilon^{23}, \varepsilon_{32}, \alpha)$ space that describes the class of system we are examining here, (i.e. coplanar, initially circular, starting at conjunction), there is a region of empirical stability defined by

$$\begin{aligned} 0 < \alpha &< \alpha_o(\varepsilon^{23}, \varepsilon_{32}) \\ 0 < \varepsilon^{23} &< 1 \qquad (2) \\ 0 < \varepsilon_{32} &< 1 \end{aligned}$$

Any system defined by parameters in this region is observed to be hierarchically stable for all time.

From the second point, there is a region of hierarchical preservation within the $(\varepsilon^{23}, \varepsilon_{32}, \alpha)$ space within which any system is guaranteed to be hierarchically preserved for all time. This region is strictly contained in the empirical stability region. This is surprising since hierarchical preservation is a necessary but not sufficient condition for hierarchical stability. It therefore seems that the analytical work of Chapter 3 is not sufficient for determining the hierarchical stability of a given system. This is particularly noticeable for retrograde systems where the differences between α_o and α_c are much greater than for the direct systems, for which $\alpha_c, \alpha_o \to 1$ as $\varepsilon^{23}, \varepsilon_{32} \to 0$.* For direct systems it is therefore seen that the analytical preservation criterion tends to the empirical stability criterion as $\varepsilon^{23}, \varepsilon_{32}$ approaches zero. However the analytical criterion never reflects the empirical criterion for retrograde systems and is in fact positively misleading, as will be discussed in Sections 6.4 and 6.5.

*Chapter 8 discusses the possibility that $\alpha_o < \alpha_c$ for direct systems with $\varepsilon^{23}, \varepsilon_{32} \lesssim 10^{-5}$.

6.2

The third point is perhaps the most important. While the region of analytical preservation is smaller for the retrograde case than it is for the direct case, the opposite is true for the region of empirical stability. It seems that for two systems with equal ε^{23} , ε_{32} , α , one retrograde and the other direct, the retrograde system will be more stable. A closer examination is necessary however.

We know that α_o (direct) < α_o (retrograde) for all pairs of ε^{23} , ε_{32} that have been studied. If we take one direct and one retrograde system with fixed ε^{23} , ε_{32} and gradually increase α then we see that both are hierarchically stable for $\alpha < \alpha_o$ (direct) $< \alpha_o$ (retrograde). When α_o(direct) $< \alpha < \alpha_o$(retrograde) then the direct system becomes unstable while the retrograde system is still stable. When $\alpha > \alpha_o$(retrograde) the situation is far less clear. Now both systems are unstable but the question still remains, which will exhibit instability sooner? Secondly, what do we mean by sooner?.

If we are merely concerned with comparing the number of synodic periods that the two systems survive for, then it is easy enough to compare the relevant graphs from Figures 4.1 and 5.1. If α is only slightly greater than α_o(retrograde), then the retrograde system will be able to "survive" for many synodic periods compared to the direct system. We have seen however that for the retrograde systems with $\varepsilon^{23} \geq \varepsilon_{32}$, the lifetimes fall off very quickly with α and may well fall below the comparable direct system lifetimes. We must also remember that commensurabilities enhance the stability of certain direct systems. These systems will almost certainly have stability lifetimes exceeding their retrograde counterparts, especially for α approaching unity.

If we are concerned with comparing absolute lifetimes, then the situation is more complicated still. Let t_R, t_D be the absolute lifetimes of the retrograde and direct system respectively. In addition let S_R, S_D be the corresponding synodic periods and $(N_s)_R$, $(N_s)_D$ be the lifetimes in units of the synodic periods as given by the fitted curve $f(\alpha)$, (Equation (1)). Thus the predicted absolute lifetimes are

$$\begin{aligned} t_R &= S_R \cdot (N_s)_R \\ t_D &= S_D \cdot (N_s)_D \end{aligned} \tag{3}$$

$S_D = 2\pi/(n_2-n_3)$ and $S_R = 2\pi/(n_2+n_3)$, where n_2 and n_3 are sidereal mean motions of the inner and outer binaries in each case. Note that n_2 is the same for the retrograde and direct system; as is n_3. Only the sign prefixing n_3 alters. Therefore

$$\frac{S_R}{S_D} = \frac{n_2-n_3}{n_2+n_3} \tag{4}$$

In a similar manner to that given in Section 5.4,

$$\begin{aligned} n_2^2\, a_2^3 &= G(m_1 + m_2) \\ n_3^2\, a_3^3 &= G(m_1 + m_2 + m_3) \end{aligned} \tag{5}$$

implies

$$\left(\frac{n_3}{n_2}\right)^2 = \alpha^3\,(1 + \mu_3) \tag{6}$$

Hence

$$\frac{n_3}{n_2} = \sqrt{\alpha^3 + \varepsilon_{32}} \tag{7}$$

Substituting Equation (7) into Equation (4), we find

$$\frac{S_R}{S_D} = \frac{1 - \sqrt{\alpha^3 + \varepsilon_{32}}}{1 + \sqrt{\alpha^3 + \varepsilon_{32}}} \tag{8}$$

Hence, from Equation (3),

$$\frac{t_R}{t_D} = \frac{1 - \sqrt{\alpha^3 + \varepsilon_{32}}}{1 + \sqrt{\alpha^3 + \varepsilon_{32}}} \cdot \frac{(f(\alpha))_R}{(f(\alpha))_D} \qquad (9)$$

where

$$f(\alpha) = \exp\left[\beta\left(\frac{1-\alpha}{\alpha - \alpha_o}\right)^{\gamma}\right] - 1 \ , \ \alpha > (\alpha_o)_R \qquad (10)$$

and the values of α_o, β, γ vary depending on whether the system is direct or retrograde.

We proceed to examine Equation (9) for $(\alpha_o)_R < \alpha < 1$, (recall $(\alpha_o)_D < (\alpha_o)_R$). As α rises towards the value $(1-\varepsilon_{32})^{1/3}$, $n_2 \to n_3$ and hence $S_D \to \infty$. $S_R, (f(\alpha))_R$, and $(f(\alpha))_D$ all remain strictly positive and finite. Hence $t_R/t_D \to 0$, i.e. the direct system is more stable than its retrograde equivalent.

As α falls towards $(\alpha_o)_R$, $(f(\alpha))_R \to \infty$. S_R, S_D and $f(\alpha)_D$ remain strictly positive and finite. Hence $t_R/t_D \to \infty$ and the retrograde system is more stable. Somewhere between $(\alpha_o)_R$ and $(1-\varepsilon_{32})^{1/3}$ there exists a value of α where $t_R/t_D = 1$, i.e. each system has the same absolute stability lifetime. This critical value of α will of course depend on the values of α_o, β, γ pertaining to each system.

To summarise, it seems that for given $\varepsilon^{23}, \varepsilon_{32}$, there exists two critical values of α; $(\alpha_o)_D$ and $(\alpha_o)_R$ where $(\alpha_o)_D < (\alpha_o)_R$. For $\alpha \in (0, (\alpha_o)_D]$, both direct and retrograde systems are hierarchically stable. For $\alpha \in ((\alpha_o)_D, (\alpha_o)_R]$, the direct systems usually exhibit instability after a finite time, while the retrograde systems are still stable for all time. For $\alpha \in ((\alpha_o)_R, 1)$, both direct and retrograde systems are unstable. Comparing equivalent direct and retrograde systems, it is seen that for α slightly greater than $(\alpha_o)_R$, the retrograde system will survive longer. However as α is increased, the

retrograde system's stability decreases faster than that of the direct system and at high α , the direct system will be the more stable, although both are likely to have very short stability lifetimes.

It is found that in general, retrograde systems are more stable than direct systems. This is contrary to the behaviour implied by the analytical criterion of Chapter 3. This has interesting implications, discussed in Sections 6.4 and 6.5.

6.3 The Role of Commensurabilities

The importance of commensurabilities in determining the stability of direct systems has already been discussed in Section 5.4, Equally clearly, they play little part in the stability of retrograde systems. The resulting behaviour was discussed in the previous section. The commensurabilities may enhance the stability of a particular direct system compared to its retrograde counterpart, whereas other direct systems characterised by neighbouring values of ε^{23} , ε_{32} , α may well be less stable than the equivalent retrograde ones.

We are therefore led to question the physical relevance of commensurabilities for retrograde systems. It has always been the view of this work that the stability of a hierarchical system is critically affected by the strength of the perturbing forces on the two binary orbits near conjunction. This view has been corroborated by Figures 4.3 which shows that for retrograde systems, the osculating elements are at local maxima near conjunction. Commensurabilities restrict the positions of conjunctions in the orbit to certain regions. In particular, certain regions may be excluded, where conjunctions would be more damaging to the stability of the system, due to a particularly close approach of m_2 to m_3, for example.

6.3

Comparing two systems with equal ε^{23} , ε_{32} , α , one retrograde, the other direct, the retrograde system will undergo many more conjunctions in a given time. We might expect that it would therefore be less stable as there would be many more close approaches of m_2 and m_3. However, the time per conjunction for which the two bodies are close is much smaller, due to their greater relative angular velocity, so that the disturbing force has less time to act on the disturbed body; the effect on the stability will be therefore less. We must ask the question: which is more disruptive, many weak perturbations or few strong ones? This question was answered in the previous section, where it was shown that the retrograde systems with many small perturbations caused by frequent conjunctions were more stable in general.

The same commensurability arises from given masses and separations of the bodies for both direct and retrograde systems. However the effect of the commensurability is completely different. Recall from Section 5.4, that a commensurability $n_2/n_3 = A_2/A_3$, $(A_2, A_3 \in \mathbb{N})$ possesses B conjunction lines (spokes). In the case of direct systems $B = A_2 - A_3$, but for retrograde systems $B = A_2 + A_3$. This means that the one-spoked commensurabilities that were the most effective in stabilizing direct systems, cannot exist for retrograde systems.

Another relevant point is that as $\alpha \to (1 - \varepsilon_{32})^{1/3}$, $n_2 \to n_3$ (Equation (7)). For direct systems this leads to systems with a one-spoked commensurability with conjunctions occurring every synodic period S, where $S \to \infty$. On the other hand, retrograde systems exhibit two-spoked commensurabilities with a synodic period that converges to a finite value. In this way the behaviour of retrograde systems converges

6.3

as $n_2 \rightarrow n_3$, while the direct systems' behaviour diverges. This may explain why it is so difficult to model the behaviour via. (α, N_s) curves of direct systems with high α values, while it is relatively easy to model the retrograde systems. Many of the strongest commensurabilities occur at $\alpha < (\alpha_o)_R$, (see Table 5.2), and are therefore acting on retrograde systems that are inherently stable in any case.

It is suggested that commensurabilities do exist for retrograde systems. Their effect on (α, N_s)curves is lessened in comparison with direct systems since, (a) the strongest commensurabilities affect already stable systems, (b) the frequency of occurrence of conjunctions converges as α increases for retrograde systems producing greater uniformity in stability.

Walker and Roy, (Paper III), in considering direct systems, assumed that the (α, N_s) data points could be modelled by a steadily decreasing curve $f(\alpha)$ (Equation (1)) with a dispersion of points from the curve caused by commensurabilities. They tacitly assumed that the commensurabilities generally enhanced the stability and therefore fitted the curve to the lowest data. The question arises, does the evidence from the retrograde systems support this assumption?

We must be careful about inferring too much about direct systems from retrograde ones, as we have just pointed out the very real physical differences in their behaviour. Having said that, the question can still be considered, although the answer is far from clear.

If we consider the retrograde systems with $\varepsilon^{23} < \varepsilon_{32}$, then we see that there is a smooth transition from lifetimes of several hundred synodic periods down to one or two. This steady decrease supports the assumption of Walker and Roy. On the other hand, for

$\varepsilon^{23} \geq \varepsilon_{32}$, the cut off between stability and immediate instability is almost instantaneous. If we were to assume this principle for direct systems, then any non-zero but finite stability lifetimes must be due wholly to commensurabilities. It seems therefore that the case is not proven.

There is a limit to which this argument can be taken. Commensurabilities are an inherent property of any system and it may be quite wrong to try and separate them out from the behaviour of a dynamical system. Perhaps the safest statement to make is to say that for direct systems, there is a spread in the (α, N_s) data due to their commensurable nature. This spread is best modelled by a probability density function based on the Gamma Distribution (Section 5.5). The curves that are fitted are useful for normalising the data and for estimating the value of α_o which denotes the bounds of hierarchical stability. We should be cautious if attaching any other physical significance to them.

Since the beginning of Chapter 5, commensurabilities have been looked on in a rather negative fashion. They have been considered as obstacles to data analysis, barring accurate predictions of stability lifetimes for direct systems. They do however enhance stability and there are a number of commensurable systems present in the Solar System (significantly more than expected, as shown by Roy and Ovenden, 1954, 1955). They are therefore of interest in their own right. Chapter 7 takes a more constructive view of them and indicates another way in which they might be detected.

6.4 Comparisons with Real Systems

In this section, the results from the studies of fictitious systems are applied to real three-body systems. The majority of real systems *considered are in fact three-body subsystems of the Solar System.* In order to apply the previous results, we need to assume that they are isolated from all perturbations apart from the mutual perturbations of the three bodies in question. They must be considered to be coplanar with near circular orbits. These assumptions are justified to a greater or lesser extent for each of the systems considered. Cases where they are not will usually be noted.

The comparisons between real systems and the numerical results are described by reference to Tables 6.1 - 6.6, at the end of this section. Each table contains a separate class of system, as will be readily seen when we examine each in turn. For each system, μ and μ_3 are calculated from the masses. These masses are taken from a number of sources, (principally Roy, 1979; Aksnes, 1985; Cambridge Atlas of Astronomy, 1985). In the case of the small satellites the masses are inferred from their sizes and densities. These densities are only estimates from spectroscopic and photometric data and therefore there is a considerable uncertainty in measuring the masses.

From μ and μ_3, the critical value α_c for hierarchical preservation may be evaluated and compared with the actual value α. If $\alpha < \alpha_c$, the system exhibits the given hierarchy for all time. The values of ϵ^{23} and ϵ_{32} may be calculated from μ, μ_3 and α. Using Tables 4.1 and 5.1, α_o can be determined using the value whose epsilon parameters agree most closely with those measured. Where the *actual epsilon parameters are less than those tabulated, a lower limit*

on α_o is given. If the actual values are greater than those tabulated, an upper limit on α_o is given. From the value of α_o, it is possible to say whether the given system is hierarchically stable or not. We now go on to study each table individually. (For a given system, a bracketed value of α_c indicates that $\alpha > \alpha_c$).

Table 6.1: Sun-Planet-Planet

Within the Solar System, the stability of the Sun and various pairs of planets are considered. One pair that is excluded is Neptune-Pluto. Since these planets have crossing orbits they are deemed to be hierarchically unstable. Their orbits are maintained because a critical argument prevents Neptune from approaching Pluto when the latter is near perihelion, (see Section 7.4). Pluto's orbit is decidedly eccentric ($e \sim 0.25$) and is inclined at $\sim 17^o$ to the Earth's orbital plane, violating our underlying assumptions. We should not attach too much significance to any results concerning this planet.

The whole class of systems in this table are characterised by the small μ, μ_3 and hence small ϵ^{23} , ϵ_{32} . This implies that both α_c and α_o are close to unity (all-systems are direct). As the actual value of α for each system is much less than both α_c and α_o, we can say with a degree of confidence that the systems considered here are strongly stable. This does not of course exclude the possibility that a planet's orbit may be disrupted by a combination of perturbations from many planets. Such a study is outside the scope of this work, although it will be discussed briefly in Chapter 8.

By studying ϵ^{23} , ϵ_{32} , we can see that Jupiter dominates in terms of its perturbations on other bodies' orbits about the Sun. Every planet receives its greatest perturbation from Jupiter - even Mercury (although Venus comes a close second).

6.4

Care must be taken in examining the stability of the systems with respect to α_o. We know the values of μ, μ_3, α and can therefore calculate ϵ^{23} and ϵ_{32}. For this pair of empirical stability parameters there is an associated value of α_o. Since $\alpha < \alpha_o$ for all the systems listed in Table 6.1, we may assume they are hierarchically stable. We may <u>not</u> assume that they remain stable if α is increased to the value α_o. This is because α_o is a function of ϵ^{23} and ϵ_{32} which in turn depend on α. As α is increased, ϵ^{23} and ϵ_{32} increase and the value of α_o decreases.

The true stability limit α_o for given μ, μ_3 is obtained by solving

$$\alpha_o = \alpha_o \left(\epsilon^{23}(\mu, \alpha_o), \epsilon_{32}(\tilde{\mu}_3, \alpha_o)\right) \qquad (11)$$

Unfortunately no analytical form for $\alpha_o = \alpha_o(\epsilon^{23}, \epsilon_{32})$ has been found. However a reasonably accurate value may be estimated from Tables 4.1 or 5.1 if desired. As an example take the Sun-Mars-Jupiter system. Here $\alpha = 0.273$ and α_o exceeds 0.95 for $\mu = 3.24 \times 10^{-7}$ and $\mu_3 = 9.55 \times 10^{-4}$. Obviously the true stability limit for these masses lies somewhere between 0.273 and 0.95. Suppose Mars is positioned nearer Jupiter (neglecting obstacles like the asteroid belt) such that $\alpha = 0.6$. Then $\epsilon^{23} = 1.17 \times 10^{-7}$ and $\epsilon_{32} = 2.06 \times 10^{-4}$ which implies that α_o has a value between 0.80 and 0.92 (Table 5.1). Thus we investigate $\alpha = 0.7$ and so on until α and α_o agree within the accuracy of measurement. In this case, $\alpha_o(\mu, \mu_3)$ is estimated as being in the range (0.8, 0.85). This is less than the original estimate of 0.95 but is still considerably higher than the actual value of α. We may therefore conclude that Mars will not be violently perturbed by Jupiter. The other systems follow a similar pattern.

6.4

Table 6.2: Planet-Satellite-Satellite

There is little difference in the sizes of the parameters in Table 6.2 compared with Table 6.1. This is not surprising since a planet and its satellites have the same hierarchical structure as the Sun and planets. The only real differences are that many satellites' orbits cross while many others show large α values which are nevertheless smaller than α_c and α_o indicating hierarchical stability. For the retrograde satellites, α usually exceeds α_c but is still comfortably less than α_o.

For the Martian system, we see that it will exhibit the same hierarchy for all time and is moreover hierarchically stable. It has been observed that the orbit of the inner satellite, Phobos, is decreasing. Phobos will crash onto the surface of Mars in about 30 million years, (hardly the mark of a stable system). This apparent contradiction is resolved in two ways. Phobos by moving towards Mars is not changing the hierarchy. The fact that $\alpha < \alpha_c$ means that the perturbations from Deimos will never allow Phobos to reverse this inward spiral and move out to cross Deimos' orbit. Since $\alpha < \alpha_o$, this means that Deimos is not responsible for the irreparable changes in Phobos' orbit but rather it is due to tidal interactions between Phobos and Mars (see Section 6.5). These tidal forces are not modelled in the classical three-body problem.

The main satellites of Jupiter can be divided naturally into four groups of four (a feature which might have appealed to Kepler). The first four are Metis, Adrastea, Amalthea and Thebe. These small satellites have such small empirical stability parameters that they must be very close before significant interaction takes place. Metis and Adrastea occupy the same orbit situated on the outer edge of Jupiter's rings. Because of the interaction between the rings and these small

6.4

satellites, any results from the point mass three-body problem must be viewed with caution.

The second group is the Galilean satellites - Io, Europa, Ganymede and Callisto. These are by far the largest of Jupiters satellites and produce the greatest disturbances on the other bodies. No single one dominates but Io being the innermost of the four has most influence on the inner four satellites, while Callisto mostly influences the outer satellites. All seem highly stable with respect to each other.

The third group comprises Leda, Himalia, Lysithea and Elara. These bodies have eccentric, inclined, orbits that cross (typically $e \sim 0.15$, $i \sim 28^{o}$). Once again these bodies are hierarchically unstable with respect to each other, therefore some other mechanism must be looked for to justify their continued proximity. Note from Table 6.2 that had Himalia and Lysithea been in circular orbits with the same semi-major axes, then they would have been hierarchically stable, being so small. This group do seem stable with respect to the other groups.

The final group is perhaps the most interesting for our purposes. It comprises Ananke, Carme, Pasiphae and Sinope. These satellites also have eccentric, inclined orbits that cross but these are retrograde, ($e \sim 0.2$, $i \sim 150^{o}$). Once again they are hierarchically unstable with respect to each other but must be maintained by some resonant behaviour. If we are to believe the analytical preservation criterion, they may also cross the orbits of the second and third groups. However by studying the values of α_o we find that they are in fact stable with respect to all other groups.

The Saturnian system is now seen to be very complex with a number of small objects in the same orbits as some of the major satellites. They are therefore by our definition hierarchically unstable, but retain their orbits by the resonances that exist between them (Aksnes, 1985). For example, Telesto and Calypso are situated at the L_4 and L_5 points of Tethys, respectively. Although there is still some discussion over the nature of Janus and Epimetheus, it now seems that these co-orbiting satellites never meet as Epimetheus describes a horse-shoe orbit in relation to Janus. The shepherd satellites S26, S27 and S28, as their names suggest, have orbits that border the rings of Saturn, with which they interact. Spirig and Waldvogel (1985) have used the three-body problem with two small masses tending to zero, to model their behaviour and predict that Janus and Epimetheus exchange orbits at close encounter to produce the horse-shoe orbit, whereas S26 and S27 do not. Such behaviour is hierarchically unstable by our definition, but seems to be repeated over astronomically long timescales.

The remaining major satellites are well spread in their orbits with α much less than both α_c and α_o, indicating high stability. Titan dominates the perturbations of the other satellites in the same manner that Jupiter dominates the Solar System.

In the case of Saturn-Titan-Hyperion, Hyperion and Titan are fairly close to α_c and α_o. Their stability is enhanced by the 4:3 commensurability in mean motions that exists between them. The outermost satellite; Phoebe, is retrograde. Once again the analytical criterion indicates the possibility that it may cross the orbits of Iapetus, Hyperion and Titan at least. However, the empirical stability limit α_o is still substantially greater than any value of α observed in connection with Phoebe, hence it is pronounced hierarchically stable.

6.4

The five principal satellites of the Uranian system are well spaced and seem hierarchically stable. No satellite dominates as four of the five have comparable masses. Only the innermost, Miranda, differs from the rest, being two orders of magnitude smaller in mass.

The system of Neptune-Triton-Nereid is retrograde with a very small value of α. Being retrograde, α_c is even smaller, hence there is no guarantee of hierarchical preservation. The value of α_o, being much greater, implies that the system is indeed stable (although with an eccentricity of 0.75, Nereid's orbit is hardly circular!) In a sense this system resembles the Martian system since the inner satellite, Triton, is in a decreasing orbit. It will eventually pass through the Roche limit of Neptune and shatter producing a ring around Neptune. When this happens, all the gas giants will have rings associated with them. It is interesting to note that Triton is the only retrograde satellite to be observed within a direct satellite's orbit, and it will disintegrate in less than 100 million years. This will mean that the only stable retrograde satellites will be the outermost satellites of Jupiter and Saturn, (see Section 5.5).

Table 6.3: Planet-Satellite-Sun

This group of systems is substantially different from the previous two. The planet and one of its satellites is the close binary being perturbed by the Sun which is the third body. For this group, $\mu \ll 1$ and $\mu_3 \gg 1$, which implies that α_c is very small. Even so, $\alpha < \alpha_c$, for most of the systems studied, the only exceptions being the third and fourth groups in the Jovian system characterised by Himalia and Pasiphae respectively.

The values of ε^{23} and ε_{32} are still much less than one, implying that α_o exceeds 0.9 in many cases. For example $\alpha_o > 0.97$, for the Saturn-Dione-Sun system. This does not mean that Dione may approach the Sun to within 3% of Saturn's orbital radius before hierarchical instability sets in. Recall that α_o decreases as α increases for constant μ, μ_3. When both $\mu, \mu_3 < 1$, α_o decreased fairly slowly with α. However, when $\mu_3 > 1$, α_o decreases more quickly and will result in a small value of α_o.

From their definition, ε^{23} and ε_{32} are a measure of the perturbations on the outer and inner binary respectively. These measures are normalised with respect to the central forces producing the Keplerian motions. Implicit in the definition of a hierarchical system, is the assumption that the perturbations on the binaries are small compared to the central forces, i.e. ε^{23}, ε_{32} are less than one, preferably much less. Any system where ε^{23} or ε_{32} approach one must be considered as a poor hierarchy for which it would be unwise to expect hierarchical stability. We may therefore define α' to be that value of α for which ε^{23} or ε_{32} equals one. For the systems here, $\varepsilon^{23} < 1$ presents no problem. ε^{23} measures the perturbation on the planet-satellite system's orbit about the Sun, due to the relative displacement of the planet and satellite. This will always be very small. The perturbation of the Sun on the planet-satellite system could be large. The hierarchy would be effectively broken if $\varepsilon_{32} = \mu_3\alpha^3$ exceeds one. In this case,

$$\alpha' = \mu_3^{-1/3}$$

which is less than one and provides a more useful stability limit than α_o as given in Tables 4.1 and 5.1. In the example of Saturn-Dione-Sun,

$\alpha' = 0.066$, a more reasonable estimate which acts as an upper limit on α for all of the Saturnian satellites. It is likely that systems with α slightly less than α' would be unstable and that the true stability limit is somewhat lower. The only way to find this limit is to perform more numerical integration experiments for constant μ,μ_3 instead of ε^{23}, ε_{32} , (Chapter 8). Note that α is much less than α' for Himalia and Pasiphae, indicating that they are possibly stable.

The largest values of ε_{32} occur for Earth-Moon and the outer satellites of Jupiter and Saturn. By far the largest value of ε^{23} occurs for Earth-Moon. We are not justified in applying our results to the Uranian system since the inclination of the satellites' orbits with respect to Uranus's orbit about the Sun is close to 90^{o}. In fact it is approximately 98^{o} therefore, if anything, the satellites should be considered as retrograde rather than direct. The results should be viewed with caution in any case.

Table 6.4: Sun-Planet-Asteroid -

Not surprisingly, Jupiter dominates the planetary perturbations on the asteroid belt ($\varepsilon_{32} \sim 10^{-4}$). Saturn is the next most important ($\varepsilon_{32} \sim 10^{-5}$). Mars ($\varepsilon^{23} \sim 10^{-7}$) is much less influential. At the outer edge of the asteroid-belt, α is close to α_c and α_o as defined by Jupiter. It is almost certainly this body which governs the extent of the belt. The values of α_c and α_o will vary only slightly with μ,μ_3 assuming that masses of the asteroids are small. Neither the complicated interactions between asteroids nor the gaps in the belt caused by resonances with Jupiter are discussed in this theory.

6.4

Table 6.5: Harrington's Experiments

In 1977, Harrington completed a number of numerical experiments of a similar nature to the ones described in this work. He was concerned with showing the existence of stable planetary orbits around binary stars. He examined three-body systems with two equal masses and one small mass. His definition of stability closely matched that of hierarchical stability. He chose systems that were both direct and retrograde. The planet was allowed to orbit one or both stellar masses in an initially circular orbit. The stellar binary could have eccentricities of 0. or 0.5 (only the zero eccentricity results are quoted in Table 6.5). The masses were as follows: Star A = Star B = 1⊙, Planet a = 10^{-3} ⊙ (approximately the mass of Jupiter), Planet b = 3.3×10^{-6} ⊙ (approximately the mass of Earth). The experiments with Star A, Star B, Planet a and Planet b involved studying the stability of each system for various separations of the bodies. The stability cut-off, analogous to α_o, was noted in each case.

Harrington chose only to examine α values whose reciprocals were integers, hence his results are rather imprecise. Broadly speaking his results agree with the results of Chapters 4 and 5. His results mostly arise from systems with high ε^{23} or ε_{32} ($\sim 10^{-2}$) which is at the limit of the present studies. They do indicate that the retrograde systems are as stable, if not more so, than their direct counterparts. He also shows that for the retrograde systems, his empirical value for the stability cut-off in α is greater than α_c. In general the value of α_o derived here is greater than Haddington's limit. There may be many explanations for this. In many cases, we can only give upper limits on α_o because of the high values of ε^{23} and ε_{32}. The actual values

Table 6.1: Systems comprising the Sun and two planets. (All systems are direct)

SYSTEM			μ	μ_3	α	α_c	α_o	ϵ^{23}	ϵ_{32}
Sun -	Mercury	- Venus	1.68 E-07	2.45 E-06	0.535	0.968	> 0.97	4.81 E-08	3.76 E-07
		- Earth		3.00 E-06	0.387	0.966	> 0.97	2.52 E-08	1.74 E-07
		- Mars		3.24 E-07	0.254	0.981	> 0.97	1.08 E-08	5.31 E-09
		- Jupiter		9.55 E-04	0.0744	0.800	> 0.97	9.29 E-10	3.93 E-07
		- Saturn		2.86 E-04	0.0405	0.859	> 0.97	2.76 E-10	1.90 E-08
		- Uranus		4.38 E-05	0.0202	0.920	> 0.97	6.84 E-11	3.60 E-10
		- Neptune		5.20 E-05	0.0129	0.916	> 0.97	2.78 E-11	1.11 E-10
		- Pluto		7.54 E-09	0.0098	0.987	> 0.97	1.62 E-11	7.13 E-15
-	Venus	- Earth	2.45 E-06	3.00 E-06	0.723	0.959	0.97	1.28 E-06	1.14 E-06
		- Mars		3.24 E-07	0.475	0.967	> 0.97	5.52 E-07	3.47 E-08
		- Jupiter		9.55 E-04	0.139	0.800	> 0.97	4.73 E-08	2.57 E-06
		- Saturn		2.86 E-04	0.0757	0.858	> 0.97	1.40 E-08	1.24 E-07
		- Uranus		4.38 E-05	0.0377	0.919	> 0.97	3.48 E-09	2.35 E-09
		- Neptune		5.20 E-05	0.0241	0.915	> 0.97	1.42 E-09	7.24 E-10
		- Pluto		7.54 E-09	0.0183	0 968	> 0.97	8.24 E-10	4.62 E-14
-	Earth	- Mars	3.00 E-06	3.24 E-07	0.656	0.965	> 0.97	1.29 E-06	9.16 E-08
		- Jupiter		9.55 E-04	0.192	0.800	> 0.97	1.11 E-07	6.78 E-06
		- Saturn		2.86 E-04	0.105	0.858	> 0.97	3.30 E-08	3.29 E-07
		- Uranus		4.38 E-05	0.0521	0.919	> 0.97	8.16 E-09	6.21 E-09
		- Neptune		5.20 E-05	0.0333	0.914	> 0.97	3.32 E-09	1.91 E-09
		- Pluto		7.54 E-09	0.0254	0.966	> 0.97	1.93 E-09	1.24 E-13

Table 6.1: (continued)

SYSTEM			μ	μ_3	α	α_c	α_o	ε^{23}	ε_{32}
Sun - Mars	-	Jupiter	3.24 E-07	9.55 E-04	0.273	0.800	> 0.95	2.41 E-08	1.94 E-05
	-	Saturn		2.86 E-04	0.160	0.859	> 0.97	8.25 E-09	1.16 E-06
	-	Uranus		4.38 E-05	0.0794	0.920	> 0.97	2.04 E-09	2.20 E-08
	-	Neptune		5.20 E-05	0.0507	0.916	> 0.97	8.32 E-10	6.77 E-09
	-	Pluto		7.54 E-09	0.0386	0.984	> 0.97	4.84 E-10	4.34 E-13
- Jupiter	-	Saturn	9.54 E-04	2.86 E-04	0.545	0.782	0.91	2.84 E-04	4.64 E-05
	-	Uranus		4.38 E-05	0.271	0.794	0.91	7.02 E-05	8.75 E-07
	-	Neptune		5.20 E-05	0.173	0.794	0.95	2.86 E-05	2.69 E-07
	-	Pluto		7.54 E-09	0.132	0.797	> 0.95	1.65 E-05	1.73 E-11
- Saturn	-	Uranus	2.86 E-04	4.38 E-05	0.497	0.851	0.91	7.07 E-05	5.39 E-06
	-	Neptune		5.20 E-05	0.317	0.850	0.95	2.88 E-05	1.66 E-06
	-	Pluto		7.54 E-09	0.242	0.857	> 0.95	1.67 E-05	1.07 E-10
- Uranus	-	Neptune	4.38 E-05	5.20 E-05	0.638	0.898	0.94	1.78 E-05	1.35 E-05
	-	Pluto		7.54 E-09	0.486	0.920	> 0.95	1.03 E-05	8.66 E-10

Table 6.2: Systems comprising one of the planets and two of its satellites

SYSTEM			Direct/ Retrograde	μ	μ_3	α	α_c	α_o	ε^{23}	ε_{32}
Mars	- Phobos	- Deimos	D	1.26 E-07	4.33 E-08	0.400	0.987	> 0.97	2.02 E-08	2.77 E-09
Jupiter	- Metis	- Amalthea	D	4.4 E-11	4.74 E-09	0.707	0.996	> 0.97	2.20 E-11	1.68 E-09
	Adrastea	- Io	D		4.69 E-05	0.303	0.919	> 0.97	4.04 E-12	1.30 E-06
		- Europa	D		2.56 E-05	0.191	0.933	> 0.97	1.61 E-12	1.78 E-07
		- Ganymede	D		7.84 E-05	0.120	0.905	> 0.97	6.34 E-13	1.35 E-07
		- Callisto	D		5.68 E-05	0.068	0.914	> 0.97	2.03 E-13	1.79 E-08
	- Amalthea	- Thebe	D	4.74 E-09	3.5 E-10	0.819	0.996	> 0.97	3.18 E-09	1.92 E-10
		- Io	D		4.69 E-05	0.429	0.919	> 0.97	8.72 E-10	3.70 E-06
		- Europa	D		2.56 E-05	0.270	0.933	> 0.97	3.46 E-10	5.04 E-07
		- Ganymede	D		7.84 E-05	0.169	0.905	> 0.97	1.35 E-10	3.78 E-07
		- Callisto	D		5.68 E-05	0.096	0.914	> 0.97	4.37 E-11	5.03 E-08
	- Thebe	- Io	D	3.5 E-10	4.69 E-05	0.524	0.919	> 0.95	9.61 E-11	6.75 E-06
		- Europa	D		2.56 E-05	0.329	0.933	> 0.97	3.79 E-11	9.12 E-07
		- Ganymede	D		7.84 E-05	0.207	0.905	> 0.97	1.50 E-11	6.95 E-07
		- Callisto	D		5.68 E-05	0.117	0.914	> 0.97	4.79 E-12	9.10 E-08
	- Io	- Europa	D	4.69 E-05	2.56 E-05	0.629	0.906	0.94	1.86 E-05	6.37 E-06
		- Ganymede	D		7.84 E-05	0.394	0.889	0.95	7.28 E-06	4.80 E-06
		- Callisto	D		5.68 E-05	0.224	0.896	> 0.97	2.35 E-06	6.38 E-07
	- Europa	- Ganymede	D	2.56 E-05	7.84 E-05	0.627	0.896	0.94	1.01 E-05	1.93 E-05
		- Callisto	D		5.68 E-05	0.355	0.903	0.97	3.23 E-06	2.54 E-06
	- Ganymede	- Callisto	D	7.84 E-05	5.68 E-05	0.567	0.886	0.94	2.52 E-05	1.04 E-05
		- Leda	D		6.8 E-13	0.096	0.904	> 0.97	7.23 E-07	6.02 E-16
		- Ananke	R		5.3 E-12	0.052	(1E-15)	> 0.997	2.12 E-07	9.56 E-17
	- Callisto	- Leda	D	5.68 E-05	6.8 E-13	0.170	0.913	> 0.97	1.64 E-06	3.34 E-15
		- Ananke	R		5.3 E-12	0.091	(2E-15)	> 0.997	4.70 E-07	5.12 E-16
	- Himalia	- Lysithea	D	4.0 E-09	5.3 E-12	0.980	0.996	(> 0.97)	3.84 E-09	4.99 E-12
		- Ananke	R		5.3 E-12	0.554	(4E-07)	> 0.997	1.23 E-09	1.16 E-13
	- Carme	- Pasiphae	D	1.8 E-11	4.4 E-11	0.959	0.999	> 0.97	1.66 E-10	3.88 E-11

Table 6.2: (continued)

SYSTEM		Direct/ Retrograde	μ	μ_3	α	α_c	α_o	ε^{23}	ε_{32}
Saturn - S26, S27, S28	- Janus	D	1.0 E-09	6.5 E-09	0.920	0.995	> 0.97	8.46 E-10	5.06 E-09
	- Mimas	D		6.68 E-08	0.749	0.990	> 0.97	5.61 E-10	2.81 E-08
	- Enceladus	D		1.46 E-07	0.586	0.987	> 0.97	3.43 E-10	2.94 E-08
	- Tethys	D		1.32 E-06	0.472	0.974	> 0.97	2.23 E-10	1.39 E-07
	- Dione	D		1.55 E-06	0.370	0.973	> 0.97	1.37 E-10	7.85 E-07
	- Rhea	D		3.87 E-06	0.264	0.963	> 0.97	6.97 E-11	7.12 E-08
	- Titan	D		2.37 E-04	0.114	0.866	> 0.97	1.30 E-11	3.51 E-07
- Janus	- Mimas	D	6.5 E-09	6.68 E-08	0.814	0.990	> 0.97	4.31 E-09	3.60 E-08
	- Enceladus	D		1.46 E-07	0.637	0.987	> 0.97	2.64 E-09	3.77 E-08
	- Tethys	D		1.32 E-06	0.514	0.974	> 0.97	1.72 E-09	1.79 E-07
	- Dione	D		1.55 E-06	0.402	0.973	> 0.97	1.05 E-09	1.01 E-07
	- Rhea	D		3.87 E-06	0.287	0.963	> 0.97	5.35 E-10	9.15 E-08
	- Titan	D		2.37 E-04	0.124	0.866	> 0.97	9.99 E-11	4.52 E-07
- Mimas	- Enceladus	D	6.68 E-08	1.46 E-07	0.782	0.986	> 0.97	4.08 E-08	6.98 E-08
	- Tethys	D		1.32 E-06	0.630	0.974	> 0.97	2.65 E-08	3.30 E-07
	- Dione	D		1.55 E-06	0.493	0.972	> 0.97	1.62 E-08	1.86 E-07
	- Rhea	D		3.87 E-06	0.353	0.963	> 0.97	8.32 E-09	1.70 E-07
	- Titan	D		2.37 E-04	0.152	0.866	> 0.97	1.54 E-09	8.32 E-07
- Enceladus	- Tethys	D	1.46 E-07	1.32 E-06	0.807	0.973	> 0.97	9.51 E-08	6.94 E-07
	- Dione	D		1.55 E-06	0.631	0.972	> 0.97	5.81 E-08	3.89 E-07
	- Rhea	D		3.87 E-06	0.452	0.963	> 0.97	2.98 E-08	3.57 E-07
	- Titan	D		2.37 E-04	0.195	0.866	> 0.97	5.55 E-09	1.76 E-06
- Tethys	- Dione	D	1.32 E-06	1.55 E-06	0.782	0.967	> 0.97	8.07 E-07	7.41 E-07
	- Rhea	D		3.87 E-06	0.560	0.960	> 0.97	4.14 E-07	6.80 E-07
	- Titan	D		2.37 E-04	0.241	0.866	> 0.97	7.67 E-08	3.32 E-06
- Dione	- Rhea	D	1.55 E-06	3.87 E-06	0.715	0.959	> 0.97	7.92 E-07	1.41 E-06
	- Titan	D		2.37 E-04	0.309	0.866	> 0.95	1.48 E-07	6.99 E-06
- Rhea	- Titan	D	3.87 E-06	2.37 E-04	0.431	0.866	> 0.95	7.19 E-07	1.90 E-05

Table 6.2: (continued)

SYSTEM			Direct/ Retrograde	μ	μ_3	α	α_c	α_o	ε^{23}	ε_{32}
Saturn - Titan	-	Hyperion	D	2.37 E-04	2.64 E-08	0.825	0.865	> 0.91	1.61 E-04	1.48 E-08
	-	Iapetus	D		2.51 E-06	0.343	0.864	> 0.95	2.79 E-05	1.01 E-07
	-	Phoebe	R		1.18 E-08	0.094	(6E-10)	> 0.997	2.09 E-06	9.80 E-12
- Hyperion	-	Iapetus	D	2.64 E-08	2.51 E-06	0.416	0.968	> 0.97	4.57 E-09	1.81 E-07
	-	Phoebe	R		1.18 E-08	0.114	(0.027)	> 0.997	3.43 E-10	1.75 E-11
- Iapetus	-	Phoebe	R	2.51 E-06	1.18 E-08	0.275	(5E-06)	> 0.997	1.90 E-07	2.45 E-10
Uranus - Miranda	-	Ariel	D	3.93 E-07	2.63 E-05	0.677	0.932	> 0.95	1.80 E-07	8.16 E-06
	-	Umbriel	D		1.60 E-05	0.487	0.942	> 0.97	9.32 E-08	1.85 E-06
	-	Titania	D		4.91 E-05	0.297	0.917	> 0.97	3.47 E-08	1.29 E-06
	-	Oberon	D		5.10 E-05	0.222	0.916	> 0.97	1.94 E-08	5.58 E-07
- Ariel	-	Umbriel	D	2.63 E-05	1.60 E-05	0.719	0.921	0.94	1.36 E-05	5.95 E-06
	-	Titania	D		4.91 E-05	0.438	0.905	0.95	5.05 E-06	4.13 E-06
	-	Oberon	D		5.10 E-05	0.378	0.905	0.97	3.76 E-06	2.75 E-06
- Umbriel	-	Titania	D	1.60 E-05	4.91 E-05	0.610	0.910	0.94	5.95 E-06	1.11 E-05
	-	Oberon	D		5.10 E-05	0.456	0.909	0.97	3.33 E-06	4.84 E-06
- Titania	-	Oberon	D	4.91 E-05	5.10 E-05	0.747	0.897	0.94	2.74 E-05	2.13 E-05
Neptune - Triton	-	Nereid	R	5.30 E-04	2.20 E-06	0.064	(4E-06)	> 0.997	2.17 E-06	5.77 E-10

Table 6.3: Systems consisting of the Sun, a planet and one of its satellites

SYSTEM			Direct/ Retrograde	μ	μ_3	α	α_c	α'	ε^{23}	ε_{32}
Sun - Earth		- Moon	D	1.22 E-02	3.33 E+05	2.57 E-03	4.48 E-03	1.44 E-02	7.96 E-08	5.65 E-03
	- Mars	- Phobos	D	1.26 E-07	3.08 E+06	4.12 E-05	2.14 E-03	6.87 E-03	2.14 E-16	2.15 E-07
		- Deimos	D	4.33 E-08		1.03 E-04	2.14 E-03		4.59 E-16	3.37 E-06
	- Jupiter	- Metis	D	4.4 E-11	1.05 E+03	1.65 E-04	2.99 E-03	9.84 E-02	1.20 E-18	4.72 E-09
		- Amalthea	D	4.74 E-09		2.33 E-04	2.99 E-03		2.57 E-16	1.33 E-08
		- Thebe	D	3.5 E-10		2.84 E-04	2.99 E-03		2.82 E-17	2.41 E-08
		- Io	D	4.69 E-05		5.42 E-04	2.99 E-03		1.38 E-11	1.67 E-07
		- Europa	D	2.56 E-05		8.62 E-04	2.99 E-03		1.90 E-11	6.73 E-07
		- Ganymede	D	7.84 E-05		1.38 E-03	2.99 E-03		1.49 E-10	2.76 E-06
		- Callisto	D	5.68 E-05		2.43 E-03	2.99 E-03		3.35 E-10	1.51 E-05
		- Himalia	D	4.0 E-09		1.47 E-02	(2.99 E-03)		8.64 E-13	3.34 E-03
		- Pasiphae	R	4.4 E-11		2.99 E-02	(1.89 E-02)		3.93 E-14	2.81 E-02
	- Saturn	- S27	D	1.0 E-09	3.50 E+03	9.78 E-05	2.02 E-02	6.59 E-02	9.56 E-18	3.27 E-09
		- Janus	D	6.5 E-09		1.06 E-04	2.02 E-02		7.30 E-17	4.17 E-09
		- Mimas	D	6.68 E-08		1.30 E-04	2.02 E-02		1.13 E-15	7.69 E-09
		- Enceladus	D	1.46 E-07		1.67 E-04	2.02 E-02		4.07 E-15	1.63 E-08
		- Tethys	D	1.32 E-06		2.09 E-04	2.02 E-02		5.77 E-14	3.20 E-08
		- Dione	D	1.55 E-06		2.64 E-04	2.02 E-02		1.08 E-13	6.44 E-08
		- Rhea	D	3.87 E-06		3.70 E-04	2.02 E-02		5.30 E-13	1.77 E-07
		- Titan	D	2.37 E-04		8.57 E-04	2.02 E-02		1.74 E-10	2.20 E-06
		- Hyperion	D	2.64 E-08		1.04 E-03	2.02 E-02		2.86 E-14	3.94 E-06
		- Iapetus	D	2.51 E-06		2.50 E-03	2.02 E-02		1.57 E-11	5.47 E-05
		- Phoebe	R	1.18 E-08		9.08 E-03	1.26 E-02		9.73 E-13	2.62 E-03
	- Uranus	- Miranda	R	3.93 E-07	2.28 E+04	4.53 E-05	6.75 E-03	3.53 E-02	8.06 E-16	2.12 E-09
		- Ariel	R	2.63 E-05		6.69 E-05	6.75 E-03		1.18 E-13	6.83 E-09
		- Umbriel	R	1.60 E-05		9.31 E-05	6.75 E-03		1.39 E-13	1.84 E-08
		- Titania	R	4.91 E-05		1.53 E-04	6.75 E-03		1.15 E-12	8.17 E-08
		- Oberon	R	5.10 E-05		2.04 E-04	6.75 E-03		2.12 E-12	1.94 E-07
	- Neptune	- Triton	R	5.3 E-04	1.92 E+04	7.90 E-05	7.14 E-03	3.73 E-02	3.31 E-12	9.47 E-09
		- Nereid	D	2.2 E-06		1.24 E-03	1.15 E-02		3.38 E-12	3.66 E-05
	- Pluto	- Charon	D	0.12	1.3 E+08	2.88 E-06	8.25 E-03	1.97 E-03	8.76 E-13	3.11 E-09

Table 6.4: Systems consisting of the Sun, a planet and an asteroid within the main asteroid belt. The asteroid is assumed to have a mass of 10^{-10} solar masses and may be positioned either in the middle of the belt or at the outer edge. All systems are direct.

SYSTEM	μ	μ_3	α	α_c	α_o	ε^{23}	ε_{32}
Sun - Mars - Asteroid (Mid Belt)	3.24 E-07	1.00 E-10	0.544	0.984	> 0.97	9.60 E-08	1.61 E-11
- Asteroid - Jupiter	1.00 E-10	9.55 E-04	0.538	0.800	> 0.80	2.90 E-11	1.49 E-04
(Mid Belt) - Saturn	1.00 E-10	2.86 E-04	0.294	0.859	> 0.95	8.61 E-12	7.23 E-06
- Asteroid - Jupiter	1.00 E-10	9.55 E-04	0.665	0.800	> 0.80	4.42 E-11	2.81 E-04
(Outer Edge)- Saturn	1.00 E-10	2.86 E-04	0.363	0.859	> 0.95	1.32 E-11	1.36 E-05

Table 6.5: Fictitious systems taken from numerical experiments by Harrington, 1977. All systems comprise a double star and planet in various hierarchical arrangements.

SYSTEM	Direct/ Retrograde	μ	μ_3	α	α_c	α_o	ε_{23}	ε_{32}
Star A - Planet a - Star B	D	1.00 E-03	1	0.250	0.244	< 0.52	6.24E-05	1.56 E-02
	R			0.333	0.151	> 0.59	1.11 E-04	3.70 E-02
- Planet b - Star B	D	3.30 E-06	1	0.250	0.244	< 0.52	2.06 E-07	1.56 E-02
	R			0.250	0.151	> 0.59	2.06 E-07	1.56 E-02
- Star B - Planet a	D	0.5	5.00 E-04	0.500	0.439	< 0.63	6.25 E-02	6.25 E-05
	R			0.333	1E-06	< 0.83	2.78 E-02	1.85 E-05
- Star B - Planet b	D	0.5	1.60 E-06	0.500	0.439	< 0.63	6.25 E-02	2.00 E-07
	R			0.500	1E-11	< 0.83	6.25 E-02	2.00 E-07
Star A - Mercury - Star B	D	0	1	0.075	0.245	> 0.80	0	4.16 E-04
- Venus	D	0	1	0.143	0.245	> 0.52	0	2.92 E-03
- Earth	D	0	1	0.200	0.245	> 0.52	0	8.00 E-03
- Mars	D	0	1	0.308	0.245	< 0.52	0	2.91 E-02

Table 6.6: Real triple star systems assumed to be both direct and retrograde

	SYSTEM	Direct/ Retrograde	μ	μ_3	α	α_c	α_o	ε^{23}	ε_{32}
ε	Hyd	D	0.448	0.958	0.0524	0.263	> 0.79	6.78 E-04	1.37 E-04
		R				0.101	> 0.94		
ADS	3358	D	0.458	0.458	0.123	0.310	> 0.60	3.75 E-03	8.50 E-04
		R				0.081	> 0.83		
-30°	529	D	0.500	0.309	0.119	0.342	> 0.60	3.54 E-03	5.20 E-04
		R				0.064	> 0.83		
ζ	Cnc	D	0.470	0.963	0.111	0.266	> 0.43	3.08 E-03	1.32 E-03
		R				0.101	> 0.82		
ξ	U Maj	D	0.270	0.807	0.0806	0.257	> 0.60	1.28 E-03	4.23 E-04
		R				0.105	> 0.83		
ζ	Aqr	D	0.180	0.820	0.0769	0.252	> 0.79	8.73 E-04	3.73 E-04
		R				0.115	> 0.94		
λ	Tau	D	0.209	0.058	0.238	0.372	> 0.60	9.37 E-03	7.85 E-04
		R				0.017	> 0.83		
	Algol	D	0.186	0.372	0.0235	0.296	> 0.91	8.34 E-05	4.81 E-06
		R				0.097	> 0.97		

6.4

will be lower. Secondly, Haddington's values are very imprecise particularly at higher α as he only samples $\alpha = 1, \frac{1}{2}, \frac{1}{3}, \ldots$. Thirdly, the way he diagnoses an instability may be different from the way presented here, both methods being open to subjectivity.

Another set of experiments that he performed assumed that the mass of Jupiter equalled the mass of the Sun. He examined the stability of the inner planets, assumed to have zero mass. He found that Mercury, Venus and Earth remained stable while Mars went unstable. The results presented here agree with Haddington insofar as $\alpha < \alpha_o$ for Mercury, Venus and Earth, while no guarantee exists for Mars.

Table 6.6: Triple Star Systems

Eight real triple star systems are studied and seen to be hierarchically stable ($\alpha < \alpha_o$). The eight are considered to be both direct and retrograde for comparison. Many of the systems show $\alpha > \alpha_c$ for retrograde motion but nevertheless α is much less than α_o. On the grounds of stability, there is no reason why retrograde triple star systems should be any less common than direct systems.

6.5 Implications for the Origin of the Solar System

All the results presented so far have been based on observed data. Any conclusions can in principle be verified by numerical integration. This section is concerned with one of the most important, if not the most important, unanswered question in celestial mechanics; namely the origin of the Solar System. What implications do the results presented here have? Can the concepts of hierarchical structure and stability contribute anything to the discussion?

We can say from the previous section that under the assumptions discussed, most of the planets and their satellites exhibit hierarchical stability. There are some exceptions such as Phobos and Triton, but the vast majority give no indication that the orbits which they currently describe will alter dramatically in the future. We may infer from this that they have followed the same paths for long periods of time in the past. The question is when and how did they fall into the stable patterns that we see today? The fact that so many are stable in their orbits does not rule out the possibility that they evolved at the same time as the Sun.

Perhaps the most important result to come out of these studies, is the fact that most retrograde systems are at least as hierarchically stable as direct systems. We must therefore ask, why are there so few retrograde bodies in the Solar System? Why are they the outermost satellites, (with the exception of Triton which we know has a relatively short lifetime)? It seems that a selection effect due to hierarchical stability is not the answer. The evidence points instead to the hypothesis that this is the way they originated. The fact that the majority of bodies are orbiting in the same direction in roughly the same plane indicates the likelihood of a common origin, rather than a series of independent events. The question of what event could cause such a diversity of bodies remain unanswered.

There are two popular theories (as well as a host of less popular derivatives). The first is that a close encounter took place between the Sun and another star. Material was passed between them from which the planets and the satellites formed. This is a rather pessimistic theory as it implies that the formation of a planetary system round a

star is a chance event and we are in a privileged position. The possibilities for life elsewhere in the Galaxy are substantially reduced if we accept that planet formation is a rare event. The second theory is that the Sun, planets and satellites condensed from one gas cloud with the Sun at the centre accreting most matter. The proto-planets' orbits about the Sun gravitationally sweep up the remaining dust and hydrogen contracting to form the system we know today. It is difficult however to explain the differing compositions of the planets by this theory.

G.H.A. Cole, in a recent paper (1985), proposes a condensation model that exhibits a greater degree of consistency than the previous model, by merely taking a different perspective of the structure of the Solar System.

Cole points out the similarity in density and composition between the Sun and the gaseous major planets Jupiter, Saturn, Uranus and Neptune. Each is a fluid body whose principal chemical component is hydrogen. Jupiter in particular is almost massive enough to be a star in its own right. By adding more material of the same type, each of the gas giants could become a star like the Sun.

On the other hand, the terrestrial planets Mercury, Venus, Earth and Mars are less massive but denser than the gas giants and composed of cosmically rarer elements. They have more in common with the satellites of the major planets than with the major planets themselves.

The model that Cole proposes assumes the Solar System originated from a cloud of material which collapsed to form five initial centres, namely the Sun and major planets. Only one centre grew to sufficient size to become stellar, although Jupiter nearly did. At the same time

6.5

or possibly later, other material collapsed to form satellites around these five gaseous centres. The terrestrial planets are considered to be the Sun's satellites. The mechanism that differentiates the silicate material from the hydrogen and helium is not clearly understood and could take place either prior or as part of the accretion of the silicate bodies.

As further evidence, Cole compares the angular momenta for various components of the Solar System. For each major planet, its rotational angular momentum is much greater than the total orbital angular momentum of its satellites. However the rotational angular momentum of the Sun is less than the total orbital angular momentum of all the planets. This inconsistency is remedied in Cole's model, since the angular momentum from the terrestrial planets is less than that of the Sun.

This model is of interest from the point of view of hierarchical systems and their stability. It breaks up the Solar System into clusters of four or five bodies. On the largest scale, there are the five gaseous bodies, i.e. the Sun and major planets, arranged in a hierarchical manner, with centre of mass very close to the Sun. The Sun has four satellites; the terrestrial planets (five if we include the Moon). Cole suggests that the natural division of the Jovian satellites into four groups of four may be more than coincidence. Uranus also has five well spaced satellites. In Cole's model, he proposes that the condensation takes place in clusters of five or six bodies that can be hierarchically ordered. The model may be applied to the formation of star clusters in a similar way. If this model is valid, then it increases the importance of studying hierarchical systems of

6.5

five or six bodies. This task would be far more difficult than the study of three bodies (see Chapter 8).

It was remarked in Section 6.4 that the retrograde satellite Triton is spiralling towards Neptune. Once it has broken up, the only retrograde satellites will be the outermost ones of Jupiter and Saturn. Considerations of hierarchical stability imply that Triton should be stable, therefore we must look to other mechanisms to describe its instability. The answer lies in the concept of tidal friction.

All the studies performed in this work have assumed the bodies to be spherically symmetric (or point) masses. In reality, the bodies are of finite size and are rotating. A nearby satellite will cause a tidal bulge in a planet. The interactions of the two non-spherical masses cause exchanges in the rotational and orbital angular momenta, producing changes in the satellite's orbit. The different types of behaviour are best described by three examples: Earth-Moon, Mars-Phobos, Neptune-Triton.

In the Earth-Moon system, the Earth's rotation period is shorter than the Moon's orbital period. The Moon produces a tidal bulge on the Earth. The Earth tries to carry the bulge round at the same rotation rate but is held back by the slowly orbiting Moon. As a consequence the Earth's rotation is slowed. It loses rotational angular momentum which is passed to the Moon in the form of orbital angular momentum. The Moon recedes as a consequence and its orbital period increases. The recession of the Moon will continue until the Earth's rotational period matches the Moon's orbital period, whereupon, the solar tidal drag decreases the angular momentum of the Earth-Moon system and the

6.5

Moon will approach the Earth again.

In the case of Mars and Phobos, the difference is that Mars' rotational period is longer than Phobos' orbital period. In this case the Mars tidal bulge is pulled round by Phobos. Mars' rotational period must increase along with its rotational angular momentum. To conserve the total angular momentum of the system, Phobos must lose orbital angular momentum. It therefore spirals in and will increase its orbital period accordingly. Unlike the Earth-Moon system, no limit is reached and Phobos will eventually crash onto Mars.

The case of Neptune-Triton is slightly different again. Because Triton is retrograde, its orbital angular momentum is of opposite sign to Neptune's rotational angular momentum. Triton is moving in the opposite direction to the bulge it is producing. Like the Earth-Moon system, the rotation of Neptune is slowed and it loses angular momentum to Triton. Unlike the Moon, the additional angular momentum is of the opposite sign to Triton's angular momentum. The result is an overall decrease in the magnitude of Triton's orbital angular momentum, producing an orbit that spirals inwards, like Phobos. There is no limit to this infall and so Triton must eventually break up within Neptune's Roche limit.

It does seem that tidal friction may account for the lack of retrograde satellites close to the planets. There are still very few retrograde satellites, even at large distances, fewer than may be expected if we assume that they are as likely to occur as direct systems. Hence the arguments concerning the origin of the Solar System still stand.

If we assume that the planets and satellites were formed by related condensation mechanisms, acting on a rotating gas cloud, it is likely that all the original bodies were orbiting in the same sense. The retrograde

satellites that we see have been captured subsequently from residual debris or orbit crossing asteroid streams such as the Apollo-Amor group. We know from our studies, that retrograde bodies are more likely to be captured than direct systems, especially at large distances where the Sun's perturbations become increasingly important.

Pluto has been studiously avoided until now. Its existence and behaviour are very difficult to explain given the present theories. Pluto is the smallest planet in the Solar System, with a composition like the terrestrial planets. Recent spectroscopic evidence shows it is covered in frozen methane like Triton. It has been suggested that Pluto orbited Neptune in the past but some catastrophic event caused it to pursue its present orbit which crosses Neptune's. At the same time Triton may have been perturbed towards Neptune. Nereid's high eccentricity may also be a consequence of the same event. There is however little evidence to support this theory and it does not properly explain Pluto's highly inclined orbit or the existence of the critical argument involving Neptune. Alternatively, Pluto may have been an asteroid which closely encountered Neptune and was locked in the critical argument that we see today. Such theories are still speculative and Pluto remains very much a mystery.

6.6 Summary

In this chapter, we have compared the behaviour of direct and retrograde systems. In Section 6.2 we concluded that retrograde systems were generally more stable than their direct counterparts. This led us to deduce that the lack of retrograde satellites and planets in the Solar System was due to tidal friction or a common mechanism for originating the bodies (Section 6.5).

6.6

The numerical results of Chapters 4 and 5 were used to compute hierarchical stability limits for real three-body systems, mostly contained in the Solar System. (The results on the Uranian system will soon need updating as more data from the Voyager flypast becomes available). By comparing the mutual distances of three bodies (α) against the empirical stability limit (α_o), we may say whether they are in a hierachically stable configuration. If they are not, we may still assume that they have been pursuing their present orbits for astronomically long time scales, but only if we look for some other dynamical mechanism that prevents them behaving in a chaotic manner. For example, we may observe critical arguments, as is the case with Neptune and Pluto, or ring interactions with nearby satellites as with the shepherd satellites of Saturn.

The stability of these systems has been described by this work only insofar as the systems are point mass, coplanar in initially circular orbits. To analyse them more accurately, there needs to be investigations into systems with small eccentricities and inclinations. This would amount to considerably more work (see Chapter 8) as would any attempt to model tidal interactions.

It was suggested in Section 6.5 that there may be many rewards in studying five and six-body systems within the Solar System. There are many complexities in studying four and more-body systems. The surface will only be skimmed in Chapters 7 and 8 and yet it is surprising how many interesting results appear from the simplest examinations.

CHAPTER 7

THE USE OF SYZYGIES IN PREDICTING STABILITY IN THE SOLAR SYSTEM

7.1 Introduction

The previous chapters have concentrated almost solely on the general three-body problem. It is however of some interest to consider the stability of four and more body systems. Whereas there is only one hierarchical arrangement of three bodies, namely $((m_1,m_2),m_3)$, there are two hierarchically distinct arrangements for four bodies. They are the simple hierarchy $(((m_1,m_2),m_3),m_4)$ and the double binary hierarchy $((m_1,m_2),(m_3,m_4))$ as was discussed in Section 2.2. For larger systems the picture becomes even more complicated. Walker (1983) described a coordinate system for a general n-body hierarchical dynamical system with generalised empirical stability parameters giving the mutual perturbations on various subsystems. Walker and Roy (1983) presented explicitly the equations of motion for both four-body hierarchies in terms of the empirical stability parameters, as well as giving the regions in the parameter space where real systems may exist stably.

Because of the many complexities involved in studying the n-body problem, it is necessary to restrict the field of study to simple, coplanar hierarchies as given in Figure 7.1. Thus it is not possible to consider multiple stellar systems such as Castor. This condition is less of a restriction within the solar system however since the orbits of planets about the Sun as well as satellites about a planet, can be considered by using a simple coplanar hierarchical model.

The importance of conjunctions in determining stability has been stressed in the preceding chapters on three-body systems. In particular, for the retrograde systems, it was particularly noticeable that the stability lifetimes were quantised into whole numbers of

7.1

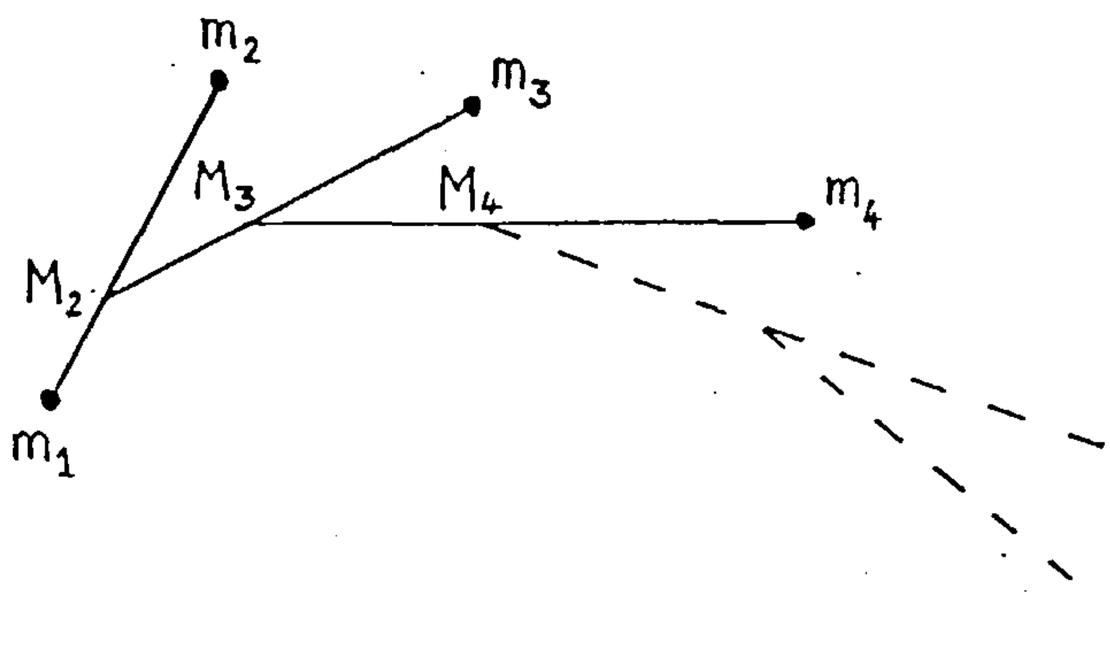


Figure 7.1 A simple n-body hierarchical system
$(\ldots(((m_1,m_2),m_3),m_4),\ldots,m_n)$
where M_i denoted the centre of mass of the subsystem
$(\ldots((m_1,m_2),m_3),\ldots,m_i)$

synodic periods. This indicated that the perturbations were greatest at a conjunction, giving the greatest opportunity for an instability to occur. This was the reason for using the synodic period as a unit of time; it gives the number of conjunctions for which a system is stable.

There is no reason to suppose that four and more bodies will behave in a radically different manner. In other words the occurrence of conjunctions is of considerable importance, in determining the stability of n-body systems. -

There is, of course, difficulty in dealing with four and more bodies, since a perfect alignment of four bodies is practically never possible. If one considers the sun and three planets, it is possible to compute a synodic period for any two of the planets about the sun. However, the probability of the third planet being exactly aligned with the other two at the instant when they are at conjunction

7.1

is zero. . Thus the definition of a synodic period as the time between successive ordered alignments of n bodies, implies an infinite period for n > 3. (The only exceptions would occur for fictitious systems, where one can choose arbitrary longitudes and mean motions, unencumbered by the problems of real data).

Walker (1980) describes one possible definition for the synodic period of a four-body system, which he further elaborates in Roy et al. (1985). His method involves computing S_{23}, S_{24}, S_{34} where S_{ij} is the synodic period of the three-body subsystem, comprising the (M_{i-1},m_i), (M_{j-1},m_j) two-body subsystems. A rough idea of the three-body subsystem that is least stable can be obtained by noting the size of the epsilon parameters; the greater the epsilon, the greater the perturbation on the relevant subsystem, and the more likely it is to exhibit an instability. Thus some synodic periods will be more important than others as a unit of time measurement. Using this basic idea, Walker has constructed a four-body synodic period S

$$\frac{1}{S} = \frac{(\varepsilon^{23} + \varepsilon_{32})/S_{23} + (\varepsilon^{24} + \varepsilon_{42})/S_{24} + (\varepsilon^{34} + \varepsilon_{43})/S_{34}}{(\varepsilon^{23} + \varepsilon_{32}) + (\varepsilon^{24} + \varepsilon_{42}) + (\varepsilon^{34} + \varepsilon_{43})}$$

where the individual three-body synodic periods have been weighted according to the relevant epsilon parameters.

This approach can be generalised to more than four bodies and gives an unambiguous unit of time for a given system. There are however two drawbacks. The first is that S is a function of the masses, as well as the mean motions, making the value more difficult to evaluate for real systems, where the masses of the bodies may not be accurately determined. The epsilon values also vary with time.

7.1

The second disadvantage is that this value of S gives no indication of the period of occurrence of alignments of the four (or more) bodies. We have already observed that a perfect alignment of more than three bodies is impossible. However if we allow alignments within a specified angular spread then it is possible to answer the question, "What is the average period of occurrence of alignments, within a certain tolerance Θ^{o}?" Such an alignment of n-bodies irrespective of conjunctions or oppositions is called a "syzygy".

A formula is derived in Section 7.2 for the average period of occurrence of syzygies from the mean motions alone. The results of numerical testing are given in Section 7.3, indicating the spread of deviations from the predicted periods for fictitious systems. These results are compared with the results for real systems in the Solar System in Section 7.4. Possible uses for this theory are discussed at the end of the chapter, where we consider not just the rotation of real bodies but also the rotation of the apse lines of orbits. This leads to a discussion into the nature of critical arguments, and mirror configurations in Section 7.6.

7.2 Predictions of the Period of Occurrence of Syzygies

Consider the sun S and p planets P_i (i=1,...,p) as in Figure 7.2. Let these planets have coplanar, circular orbits, so that their angular velocities are equal to their mean motions n_i(i=1,...p). Without loss of generality, order the planets such that

$$n_i \geq n_{i+1} \quad , \quad \forall_i = 1,2,...,p-1 \qquad (1)$$

7.2

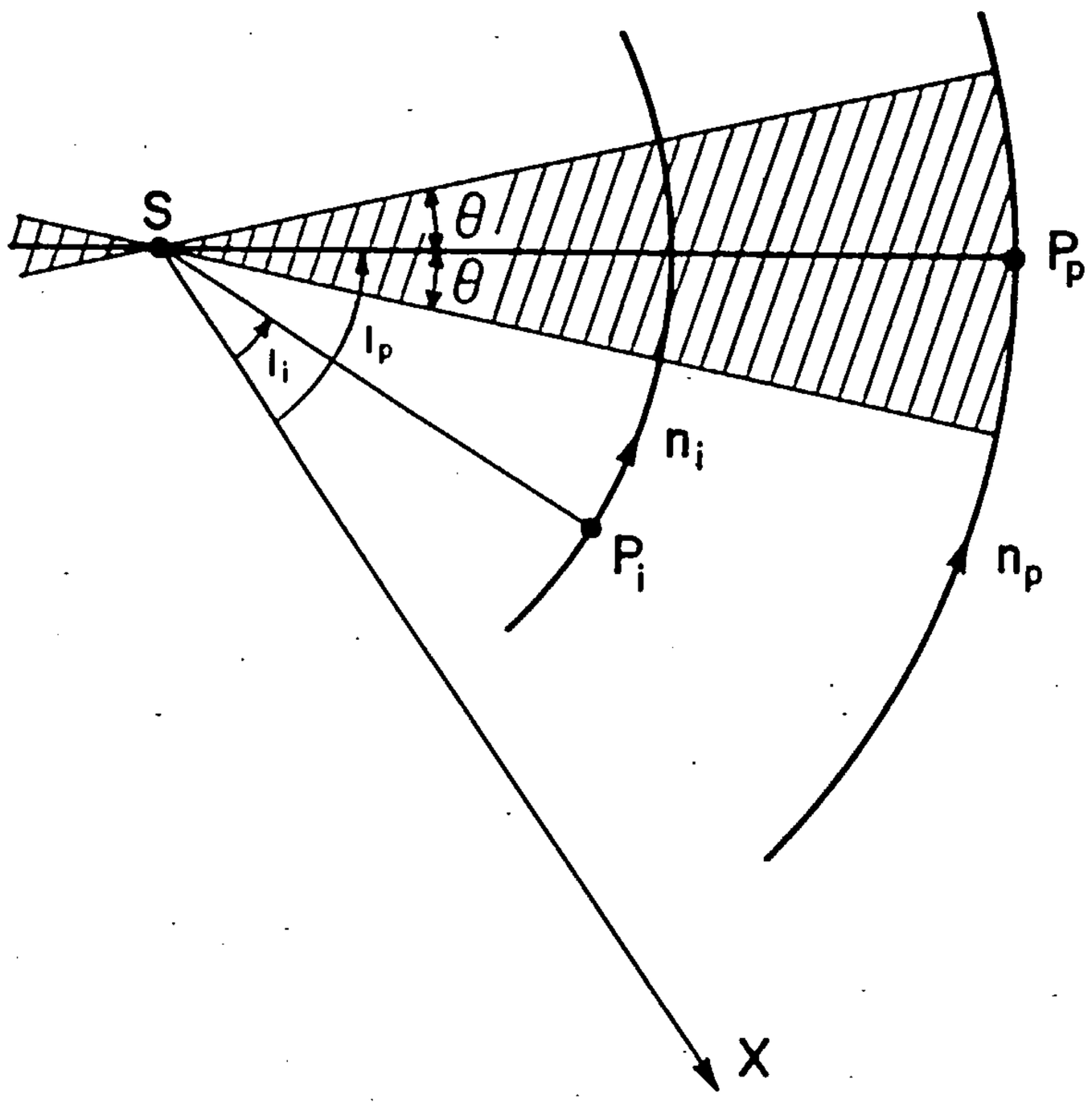


Figure 7.2. Illustration of the orbit of planet P_i with respect to planet P_p (the outermost planet). When P_i passes through the shaded region, then a syzygy of S, P_i, P_p has occurred to within a tolerance θ.

7.2

Let these orbits be unperturbed by the other planets in the system. Let $\ell_i = \angle XSP_i$ be the longitude of planet P_i with respect to fixed axis SX. (Note that a negative mean motion indicates rotation in the opposite direction from a positive mean motion). Let the required tolerance on the syzygy be Θ . This means that the system is defined to have undergone a near syzygy if

$$\left.\begin{array}{l} |\ell_i - \ell_j| \leq \Theta \\ \text{or} \\ |\ell_i - \ell_j - \pi| \leq \Theta \end{array}\right\} \forall i,j=1,2,\ldots,p \qquad (2)$$

Because it is only necessary to consider the differences in longitudes and mean motions, define

$$\left.\begin{array}{l} L_i = \ell_i - \ell_p \\ N_i = n_i - n_p \end{array}\right\} \forall i = 1,2,\ldots,p-1$$

and set the condition on a near syzygy to be (from Inequalities (2))

$$\left.\begin{array}{ll} & |L_i| \leq \Theta \\ \text{or} & |L_i - \pi| \leq \Theta \\ \underline{\text{and}} & |L_i - L_j| \leq \Theta \\ \text{or} & |L_i - L_j - \pi| \leq \Theta \end{array}\right\} \forall i,j=1,\ldots,p-1 \qquad (3)$$

Thus the system is considered in a rotating frame with angular velocity equal to n_p and reference direction along the radius vector $\overrightarrow{SP}_p$. The evolution of the system is given by a straight line in the (p-1)-dimensional space of $(L_1,L_2,\ldots,L_{p-1})$ (See Figure 7.3). This line has parametric equations

$$L_i = L_{io} + N_i t \qquad i=1,\ldots,p-1 \qquad (4)$$

7.2

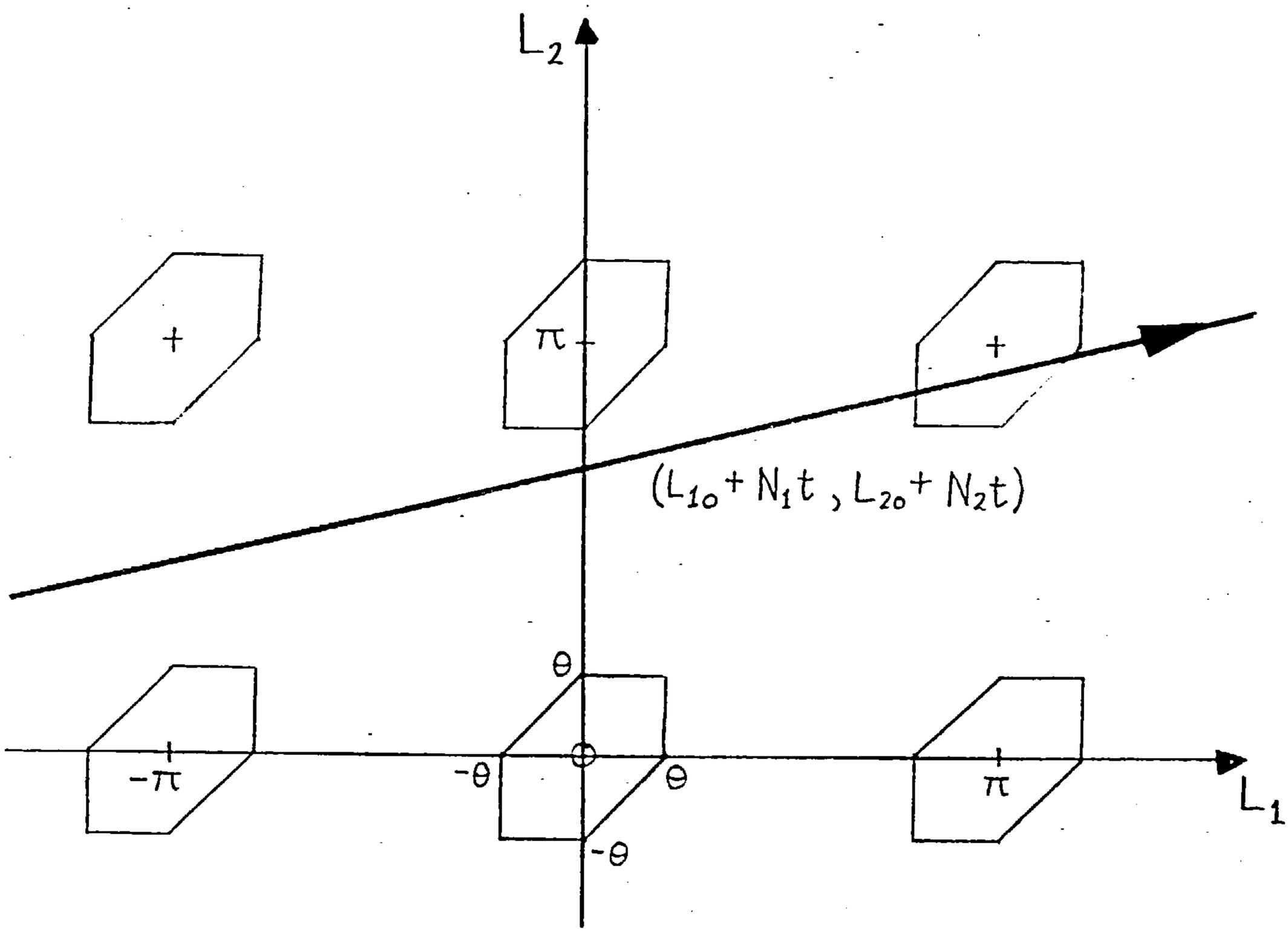


Figure 7.3: Regions of acceptable syzygies in the (L_1, L_2) plane for three planets. The evolution line of a system is shown to be at near syzygy for $L_1 \simeq L_2 \simeq \pi$.

7.2

where t is time and L_{io} is the value of L_i at $t=0$. In this configuration space there is a small region of near syzygies around each "lattice point" $(k_1\pi, k_2\pi,\ldots,k_{p-1}\pi)$, $(k_1,\ldots,k_{p-1} \in \mathbb{Z})$, with all angles measured in radians. (For example, when all the k_i's are even, the configuration is a conjunction of all the planets at the same side of the sun).

The boundaries of these regions are given by points in the (p-1)-dimensional space at which one of Inequalities (3) becomes an equality, i.e., a linear equation in L_i $(i=1,\ldots,p-1)$. Thus these boundaries are constructed from (p-2) - dimensional hyperplanes. They intersect to form a hyperpolyhedron about each lattice point. By way of example, Figure 7.5 shows such a region for p=4 (i.e. three dimensions), where each face is a region of a two-dimensional plane. The extent of the face is governed by the locus of points which satisfy Inequalities (3) while one of the inequalities is a strict equality.

For $p > 4$, the hyperpolyhedron is bounded by regions of hyperplanes in exactly the same manner. We shall describe these regions as hyperfaces and refer to the hypervolume that they occupy in (p-2) - dimensional space as "areas".

The shape and orientation of each hyperpolyhedron is the same and the size varies with the value of θ. The object is to find out how often the line (Equation (4)) passes through one of these regions.

Consider the system to be evolving by travelling along the line in $(L_1,\ldots,L_{p-1})$ space with constant velocity. The average distance between intersections with regions of syzygies is d where

$$d = (\sigma\tau)^{-1}$$

7.2

where τ is the number density of discrete regions of syzygies in the $(L_1,\ldots,L_{p-1})$ space and σ is the (p-2) - dimensional cross-section of each region in the direction of travel of the system. The direction of travel is $\underset{\sim}{N} = (N_1,N_2,\ldots,N_{p-1})$ and the average time t_p between intersections (i.e. syzygies) is given by

$$d^2 = t^2 \sum_{i=1}^{p-1} N_i^2$$

Combining this equation with the fact that $\tau = 1/\pi^{p-1}$ gives

$$t_p = \frac{\pi^{p-1}}{\sigma} \left(\sum_{i=1}^{p-1} N_i^2 \right)^{-\frac{1}{2}} \qquad (5)$$

Thus it is required to find σ. Note that all vectors are (p-1) - dimensional.

Consider one region of syzygies Ω. Suppose it is bounded by m hyperfaces with areas A_j (j=1,...,m). Associated with each hyperface is a unit normal vector $\underset{\sim}{q}_j$ (j=1,...,m). To find the cross-section it is necessary to find the projection of each hyperface along $\underset{\sim}{N}$. These projections are $(\underset{\sim}{N}.\underset{\sim}{q}_j)A_j/|\underset{\sim}{N}|$ (j=1,...,m), remembering that although $\underset{\sim}{q}_j$ are unit vectors, $\underset{\sim}{N}$ is not. The cross-section is derived by summing the projections of the hyperfaces, taking care only to sum over projections with the same sign. (If one sums over all possible faces, then one is considering hidden faces and $\sigma = 0$). Without loss of generality, assume that the summing is done over positive projections. Therefore

$$\sigma = \sum_{j=1}^{m} \max \left[0, \frac{(\underset{\sim}{N}.\underset{\sim}{q}_j)}{|\underset{\sim}{N}|} A_j \right] \qquad (6)$$

Combining Equation (6) with Equation (5) and noting that $|\underset{\sim}{N}|^2 = N_1^2 + \ldots + N_{p-1}^2$ gives

7.2

$$t_p = \pi^{p-1} \Big/ \sum_{j=1}^{m} \left[\max (0, \underset{\sim}{N}.\underset{\sim}{q}_j).A_j \right] \qquad (7)$$

Having found the relevant unit normal $\underset{\sim}{q}$ say, the corresponding area A is given by the following result.

Theorem

$$A = \int_{B_i} \frac{1}{|q_i|} \left[\prod_{\substack{j=1 \\ j \neq i}}^{p-1} dL_j \right] \qquad (8)$$

where q_i denotes the i^{th} component of unit normal vector $\underset{\sim}{q}$, assuming $q_i \neq 0$, and B_i is the projection of the relevant hyperface onto the (p-2) - dimensional subspace $L_i = 0$.

Proof for p=4

For p=4, we are considering the three-dimensional space (L_1, L_2, L_3) (see Figure 7.4). It is required to find the area A of surface Σ. Let it be projected onto the (L_1, L_2) plane as surface Σ_3. Σ can be divided into small areal elements dS, whose projections on the (L_1, L_2) plane are rectangles with sides dL_1, dL_2, as shown. Let $\underset{\sim}{q}=(q_1,q_2,q_3)$ be the unit normal to dS in either direction. Clearly $dL_1.dL_2 = dS.\ |\cos\gamma| = dS\ |q_3|$. Thus

$$A = \int_{\Sigma} dS = \int_{\Sigma_3} \frac{dL_1\ dL_2}{|q_3|}$$

where q_3 may be a function of L_1 and L_2. Care must be taken to ensure that q_3 is never zero. Similar results can be obtained by projecting Σ onto Σ_2 (on the (L_1, L_3) plane) or Σ_1 (on the (L_2, L_3) plane). Thus

$$A = \int_{\Sigma_1} \frac{1}{|q_1|}\, dL_2 dL_3 = \int_{\Sigma_2} \frac{1}{|q_2|}\, dL_1 dL_3 = \int_{\Sigma_3} \frac{1}{|q_3|}\, dL_1 dL_2 .$$

7.2

This theory can be generalised to higher dimensions, the result being Equation (8).

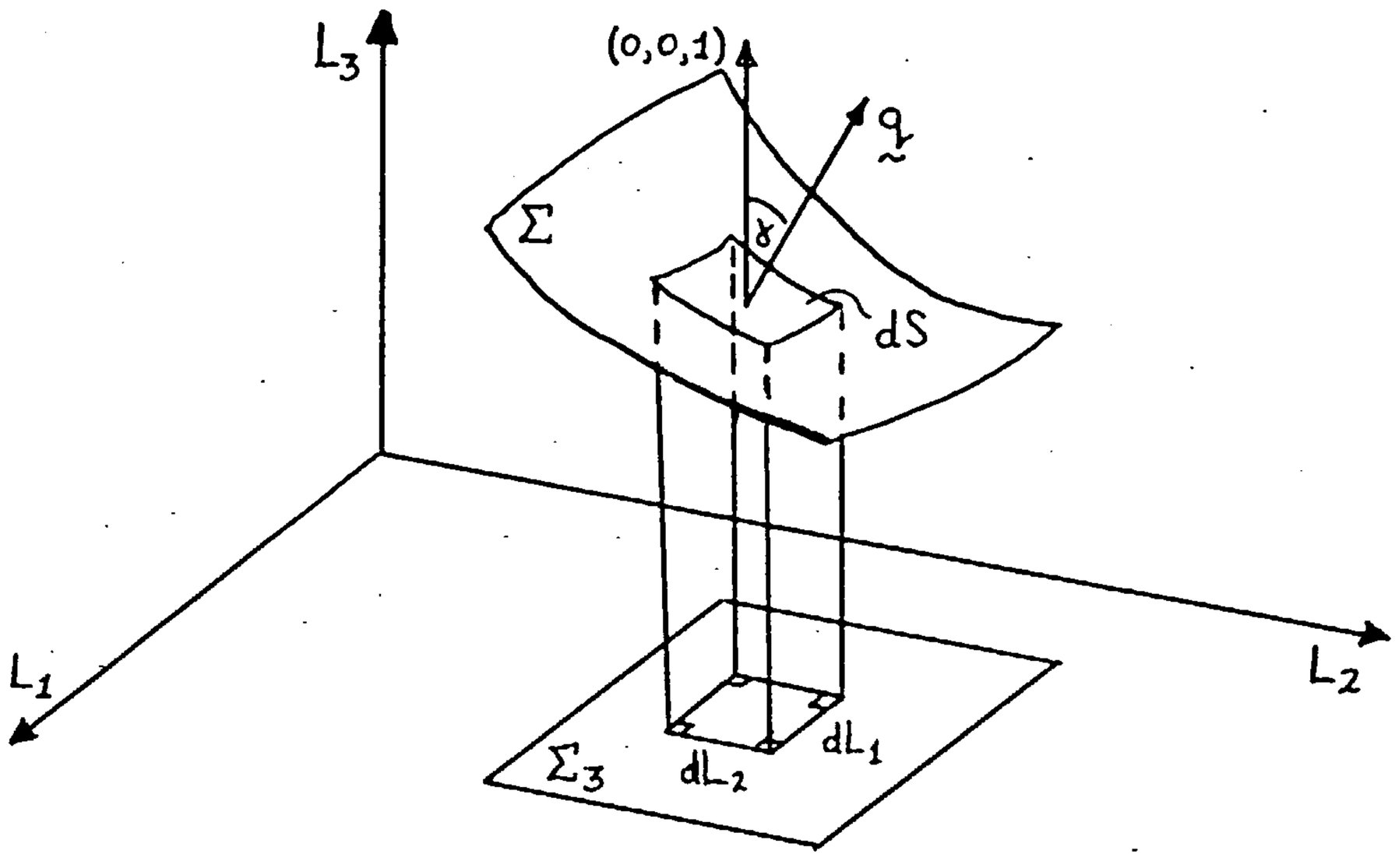


Figure 7.4 Surface Σ projected onto Σ_3 on the (L_1, L_2)-plane

It is now possible to find $\underset{\sim}{q}_j$, A_j $(j=1,\ldots,m)$. Equation (7) is independent of the initial longitudes, hence it is only necessary to examine the region Ω around the origin. The results obtained apply to regions centred at other lattice points with coordinates that are integer multiples of π. Thus from Inequalities (3), the conditions for near syzygy become

$$\left.\begin{aligned} |L_i| &\lesssim \theta \\ \text{and} \quad |L_i - L_j| &\lesssim \theta \end{aligned}\right\} \quad \forall i,j=1,\ldots,p-1 \qquad (9)$$

To evaluate $\underset{\sim}{q}_j$ and A_j it is necessary to consider the hyperfaces in three categories.

7.2

a) $L_i \geq 0, \forall i=1,2,\ldots,p-1$

Inequalities (9) become

$$\max_{i=1,\ldots,p-1} (L_i) \leq \theta$$

Let $L_k = \max (L_i)$ for some $k=1,\ldots,p-1$. The relevant boundary hyperface has equation $L_k = \theta$ and extends over the region where $L_i \leq L_k$, $\forall i=1,\ldots,p-1$. The normal vector to the hyperface $\underset{\sim}{q}=(0,\ldots,0,1,0,\ldots,0)$ where the 1 is in the k^{th} place. From Equation (8),

$$A = \int_o^{L_k} \cdots \int_o^{L_K} \prod_{\substack{j=1 \\ j\neq k}}^{p-1} dL_j$$

with $0 \leq L_j \leq L_k = \theta, \quad \forall j=1,\ldots,p-1$. Thus

$$A = \prod_{\substack{j=1 \\ j\neq k}}^{p-1} \int_o^{\theta} dL_j = \theta^{p-2}$$

Clearly there are p-1 distinct faces in this category to be considered.

b) $L_i \leq 0, \quad \forall i=1,2,\ldots,p-1$

The theory is similar to (a).

$\underset{\sim}{q} = (0,\ldots,0,-1,0,\ldots,0)$ (-1 in k^{th} place), $k=1,\ldots,p-1$

$A = \theta^{p-2}$

c) $\exists i, j=1,\ldots,p-1$ such that $L_i < 0, L_j > 0$

Inequalities (9) become

$$\max_{j=1,\ldots,p-1} (L_j) - \min_{i=1,\ldots,p-1} (L_i) \leq \theta$$

7.2

Let $L_k = \max(L_i)$, $L_m = \min(L_i)$. There are (p-1) (p-2) possible combinations of k and m. Choosing a particular pair of values, the equation of the boundary hyperface is

$$L_k - L_m = \Theta$$

The unit normal vector $\underset{\sim}{q}$ has all zero components except for the k^{th} and m^{th} components which are equal to $1/\sqrt{2}$ and $-1/\sqrt{2}$ respectively.

Projecting the hyperface onto B_k we get

$$A = \int_{B_k} \sqrt{2} \prod_{\substack{j=1 \\ j\neq k}}^{p-1} dL_j$$

$$= \sqrt{2} \int_{-\Theta}^{0} \left[\int_{L_m}^{L_k} \cdots \int_{L_m}^{L_k} \prod_{\substack{j=1 \\ j\neq k,m}}^{p-1} dL_j \right] dL_m$$

$$= \sqrt{2} \int_{-\Theta}^{0} \left[\prod_{\substack{j=1 \\ j\neq k,m}}^{p-1} \int_{L_m}^{L_m+\Theta} dL_j \right] dL_m$$

$$= \sqrt{2} \int_{-\theta}^{0} \Theta^{p-3} \, dL_m$$

$$= \sqrt{2} \; \theta^{p-2}$$

The same result can be obtained by projecting the hyperface onto B_m.

The number of hyperfaces m=p(p-1). Considerable simplification of Equation (7) is now possible. As a consequence of the ordering of the mean motions (Inequalities (1)), $N_1 \geq N_2 \geq \ldots \geq N_{p-1} \geq 0$. Although

7.2

Table 7.1 lists the area and normal of all faces, those of groups (b) and (d) are never used, being hidden. (Note that the hyperfaces always occur in parallel pairs). Thus from Table 7.1, the group (a) faces give

$$(\underset{\sim}{N}.\underset{\sim}{q})A = N_i\ \theta^{p-2} \qquad i=1,\ldots,p-1$$

and the group (c) faces give

$$(\underset{\sim}{N}.\underset{\sim}{q})A = (N_i - N_j)\theta^{p-2} \qquad j=2,\ldots,p-1;\quad i=1,\ldots,j-1$$

Therefore Equation (7) becomes

$$t_p = \frac{\pi^{p-1}}{\theta^{p-2}} \Bigg/ \left[\sum_{j=1}^{p-1} N_j + \sum_{j=2}^{p-1} \sum_{i=1}^{j-1} (N_i - N_j) \right]$$

	Area	Normal	
(a)	θ^{p-2}	$(0,\ldots,0,1,0,\ldots,0)$ (i^{th} place)	$i=1,\ldots,p-1$
(b)	θ^{p-2}	$(0,\ldots,0,-1,0,\ldots,0)$ (i^{th} place)	$i=1,\ldots,p-1$
(c)	$\sqrt{2}\theta^{p-2}$	$(0,\ldots,1/\sqrt{2},\ldots,-1/\sqrt{2},\ldots,0)$ (i^{th} place) (j^{th} place)	$j=2,\ldots,p-1$ $i=1,\ldots,j-1$
(d)	$\sqrt{2}\theta^{p-2}$	$(0,\ldots,-1/\sqrt{2},\ldots,1/\sqrt{2},\ldots,0)$ (i^{th} place) (j^{th} place)	$j=2,\ldots,p-1$ $i=1,\ldots,j-1$

Table 7.1 Area and unit normal vectors for all (p-2)-dimensional hyperfaces around a region of acceptable syzygies in the (p-1)-dimensional space. θ is the maximum angular deviation from an exact syzygy that is allowable.

7.2

Theorem

$$\sum_{j=1}^{p-1} N_j + \sum_{j=2}^{p-1}\sum_{i=1}^{j-1} (N_i - N_j) = \sum_{j=1}^{p-1} N_j(p+1-2j) \quad (p=3,4,\ldots)$$

$$= \sum_{j=1}^{p} n_j(p+1-2j)$$

Proof (by induction)

Let $p=3$.

$LHS = N_1 + N_2 + (N_1 - N_2) = 2N_1$

$RHS = 2N_1 + 0.N_2 \quad = 2N_1 = LHS.$

Assume the proposition is true for some $p \geq 3$. To complete the proof, it is required to show that if the proposition is true for p then it is also true for p+1.

$$\sum_{j=1}^{p} N_j + \sum_{j=2}^{p}\sum_{i=1}^{j-1} (N_i - N_j)$$

$$= \sum_{j=1}^{p-1} N_j + \sum_{j=2}^{p-1}\sum_{i=1}^{j-1} (N_i - N_j) + N_p + \sum_{i=1}^{p-1} (N_i - N_p)$$

$$= \sum_{j=1}^{p-1} N_j(p+1-2j) + N_p + \sum_{j=1}^{p-1} (N_j - N_p)$$, assuming the proposition holds for p.

$$= \sum_{j=1}^{p-1} (N_j(p+1-2j) + N_j) - (p-1)N_p + N_p$$

$$= \sum_{j=1}^{p-1} N_j((p+1) + 1-2j) - (p-2) N_p$$

$$= \sum_{j=1}^{p} N_j((p+1) + 1 - 2j)$$

7.2

Thus the proposition holding for p, implies that it holds for p+1, as required.

$$N_j = n_j - n_p .$$

Thus

$$\sum_{j=1}^{p-1} N_j(p+1-2j)$$

$$= \sum_{j=1}^{p-1} (n_j - n_p)(p+1-2j)$$

$$= \sum_{j=1}^{p-1} n_j(p+1-2j) - n_p \sum_{j=1}^{p-1} (p+1-2j)$$

$$= \sum_{j=1}^{p-1} n_j(p+1-2j) - n_p\left[(p-1)(p+1) - 2 \sum_{j=1}^{p-1} j \right]$$

$$= \sum_{j=1}^{p-1} n_j(p+1-2j) - n_p\left[(p-1)(p+1) - (p-1)p \right]$$

$$= \sum_{j=1}^{p-1} n_j(p+1-2j) - (p-1)n_p$$

$$= \sum_{j=1}^{p} n_j(p+1-2j)$$

Q.E.D.

From this theorem we obtain the final results, namely

$$t_p = \frac{\pi^{p-1}}{\theta^{p-2}} \Big/ \sum_{j=1}^{p-1} N_j(p+1-2j) \qquad (10)$$

$$\doteq \frac{\pi^{p-1}}{\theta^{p-2}} \Big/ \sum_{j=1}^{p} n_j(p+1-2j)$$

It is instructive to derive explicitly the syzygy periods for small values of p.

7.2

p=2

$$t_2 = \frac{\pi}{N_1}$$

which is half a classical synodic period, or the time between conjunction and opposition.

p=3

$$t_3 = \frac{\pi^2}{\Theta} \cdot \frac{1}{2N_1}$$

The region Ω in the (L_1,L_2) space is given in Figure 7.3.

p=4

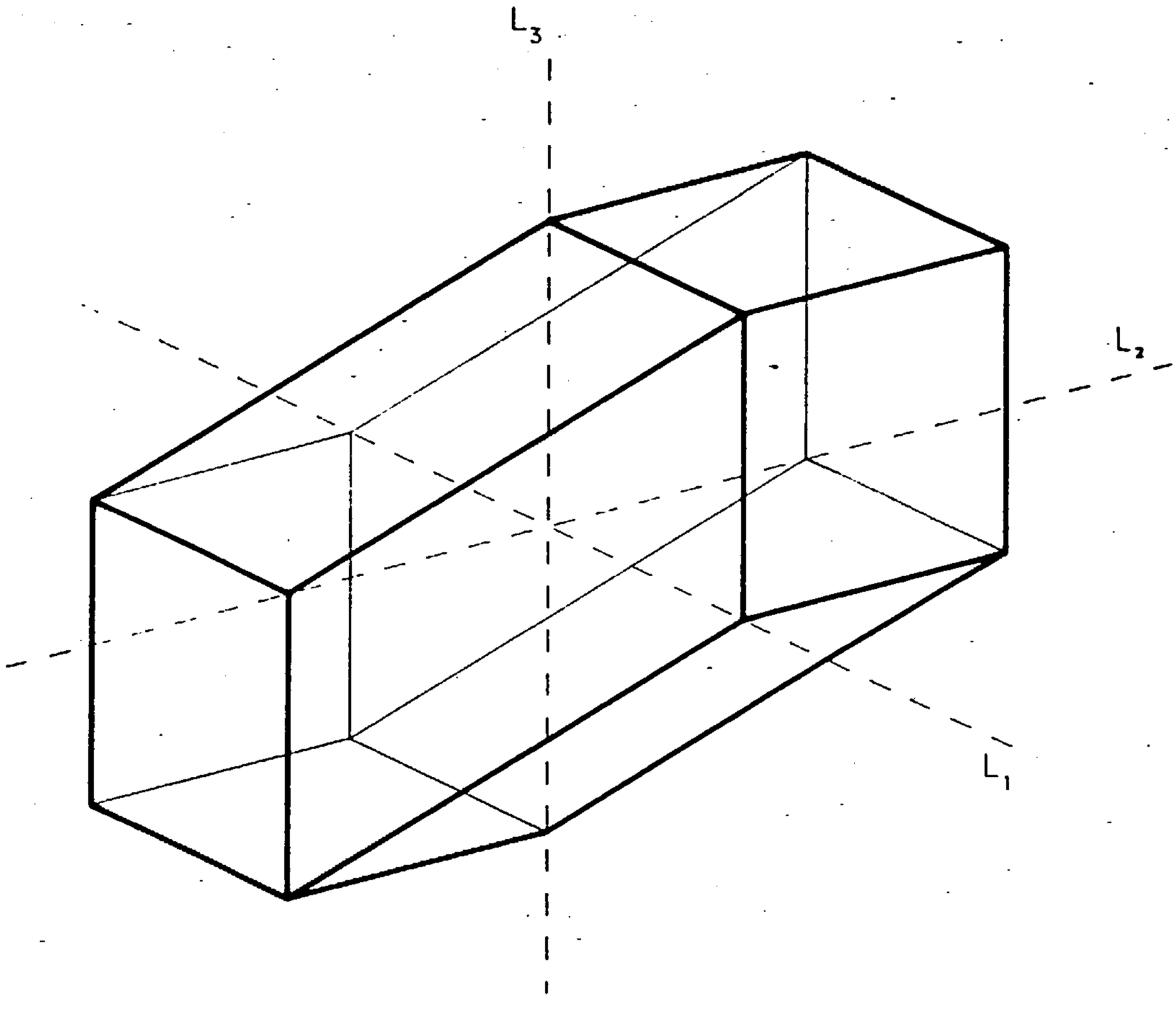
$$t_4 = \frac{\pi^3}{\Theta^2} \cdot \frac{1}{3N_1 + N_2 - N_3}$$

The region in the (L_1,L_2,L_3) space is described in Figure 7.5.

It should be noted that when there is an odd number of planets, t_p is independent of the mean motion of the middle planet.

Equation (10) gives the average period of syzygy occurrence for systems comprising of a sun and p planets with mean motions, relative to the slowest (or fastest retrograde), equal to N_j, j=1,...,p-1. If it is desirable to find the average period of occurrence of conjunctions alone, rather than syzygies, then it is necessary to multiply the time by 2^{p-1}. This may be of use if we want a measure of the synodic period of a dynamical system. In the remainder of this chapter however, we shall be interested in syzygies rather than conjunctions.

The averaging is done over all possible initial positions, therefore the actual frequency of occurrence of syzygies may vary a great deal, depending on the initial configuration and how commensurable the mean motions are. To take an extreme example if all the planets have identical



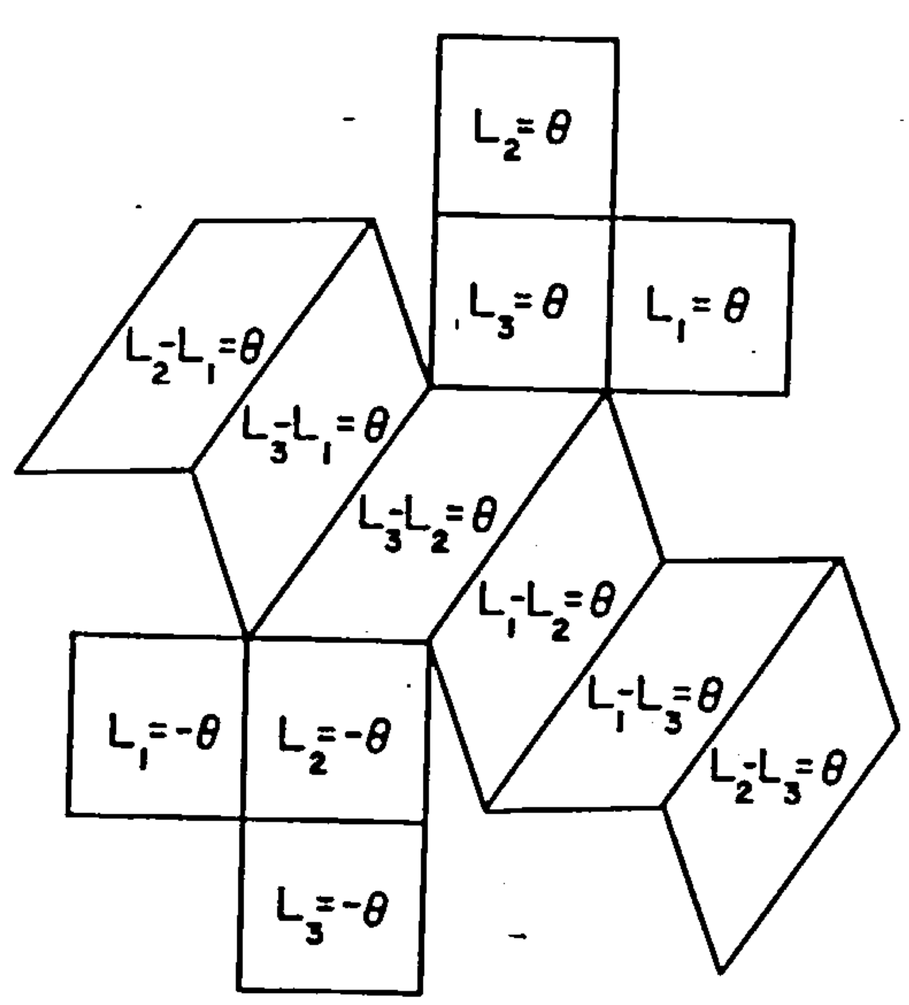


Figure 7.5: A region of acceptable syzygies in the (L_1, L_2, L_3) space for four planets.

(i) The region is a three-dimensional polyhedron centred on the origin.

(ii) A net of the polyhedron is presented with the equations of the faces.

7.2

mean motions then $L_i = L_{io}$ $\forall i=1,\ldots,p-1$ for all time, thus either the system is at permanent syzygy or it is not, depending on the initial conditions. The importance of this point is more fully discussed in Section 7.5, but it is of interest to see the actual spread of deviations from predictions. This is done by numerical experiment in the next section.

7.3 Numerical Experiments for Fictitious Systems

In order to test the accuracy of the theory, many numerical experiments have been performed by investigating fictitious systems of p planets in circular orbits about a sun. Note that the distances and masses of the planets are not considered in these experiments, only their mean motions n_i and initial longitudes ℓ_{io}. The longitudes ℓ_i at any subsequent time are given by

$$\ell_i = \ell_{io} + n_i t$$

The mean motions and initial longitudes are chosen randomly by equating $n=10^{-3r}$, $\ell_o = 360^o r$, where r is a random number between 0 and 1. The mean motions are chosen in this way to simulate the range of mean motions present in the solar system. The mean motions are ordered in decreasing size, (i.e. $n_1 > n_2 > \ldots > n_p$). The time t_p for a syzygy of p planets to accuracy Θ is calculated by Equation (10). The computer then counts the number of syzygies encountered by the system on a timescale equal to $50t_p$. Thus the predicted number of syzygies is always 50. Figures 7.6 show histograms of the actual counts observed for three, four and five-planet systems.

It is seen that the distributions are non-normal and obviously skew. One can apply normal statistics however by omitting the data from the wings. This is done by excluding all counts that lie more than 3σ away from the observed mean, σ being the observed standard deviation. The mean and standard deviation are recalculated from the truncated data and the procedure is repeated until the standard deviation converges to a constant value. The resulting data is thus approximated to a normal distribution.

7.3

The reliability of the theory can now be tested, by comparing the observed mean with the actual mean, using the Student t-test. If $\bar{X}$ denotes the observed mean, the expected mean, μ, being equal to 50, then the t-statistic is calculated as

$$t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

where n is the sample size. The 95% level for a two-tailed test is 1.98 for a sample of approximately 100 points. This means that if we accept that the theory is correct, then the value of t calculated from the observations, will lie outside the range (-1.98, 1.98) with a probability of only 5%. Thus, if this happens, it is sufficiently unlikely to raise doubts about the theory's validity.

As has been implied already, for samples of 100 or more, the resulting distribution has to be approximately normal for the t-test to be legitimately applied. The normality of the data can be tested by considering the skewness S_k and kurtosis K of the distribution. These are measured by calculating

$$S_k = m_3/m_2^{3/2}$$

$$K = m_4/m_2^2$$

where $$m_r = \left(\sum_{i=1}^{n} (x_i - \bar{X})^r\right)/n \quad , \qquad r = 2,3,4$$

The skewness is a measure of the symmetry of the distribution, and the kurtosis is a measure of the thickness of the distribution compared to the wings. The 95% limits within which S_k,K should lie, for different sample sizes, are given in Appendix A.

A summary of the results of the numerical experiments is given in Table 7.2, for various values of p and Θ. An underlined value of t, S_k, K indicates failure at the 95% level.

Figs. 7.6	P	θ^o	Fine Mesh Size	Total Sample Size	Useful Sample Size	$\bar{X}$	σ	t	S_k	K
a	3	10	11	100	92	50.13	1.93	0.647	-0.222	3.752
b	3	5	11	200	184	49.99	2.42	-0.061	-0.107	6.291
c	3	1	11	118	111	50.23	1.68	1.465	-0.131	5.234
d	4	10	11	150	141	50.02	2.86	0.088	0.200	3.143
e	4	5	11	150	142	49.82	2.68	-0.784	0.042	3.617
f	4	1	11	116	110	49.46	3.60	-1.564	0.199	3.741
g	5	5	11	124	118	48.95	3.93	-2.907	0.259	3.382
h	5	10	7	140	131	48.10	3.76	-5.783	0.021	3.474
i	5	10	11	150	143	49.11	4.07	-2.609	0.191	4.837
j	5	10	51	137	133	48.98	3.70	-3.160	-0.424	3.780
k	5	20	161	173	165	49.92	4.02	-0.271	-0.043	3.702

Table 7.2 Summary of the results from numerical experiments on randomly chosen fictitious systems. The predicted mean is 50. An underlined t value indicates a lack of agreement between the observed and predicted mean. An underlined skewness or kurtosis value indicates a lack of normality in the distribution.

(a)

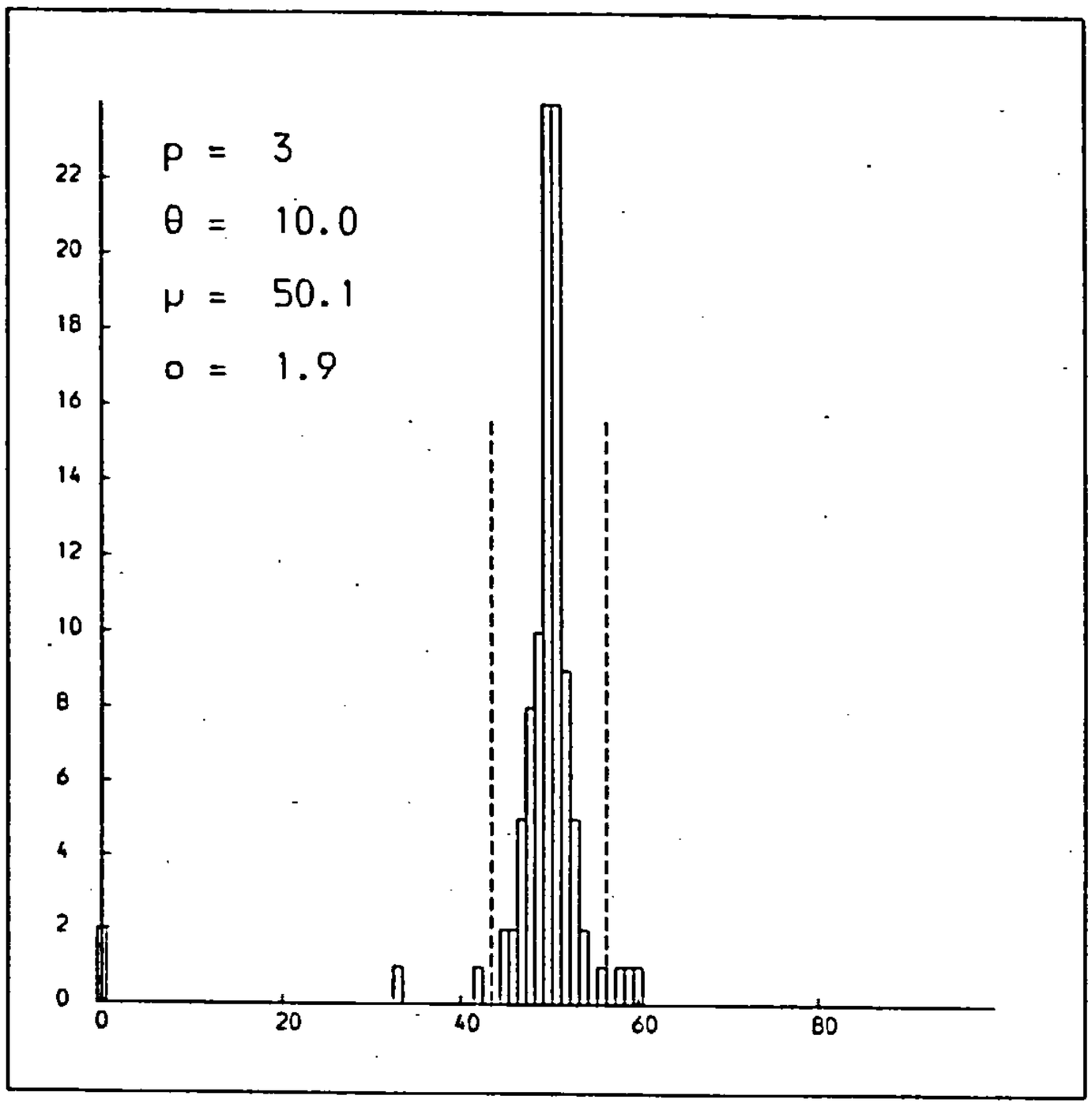
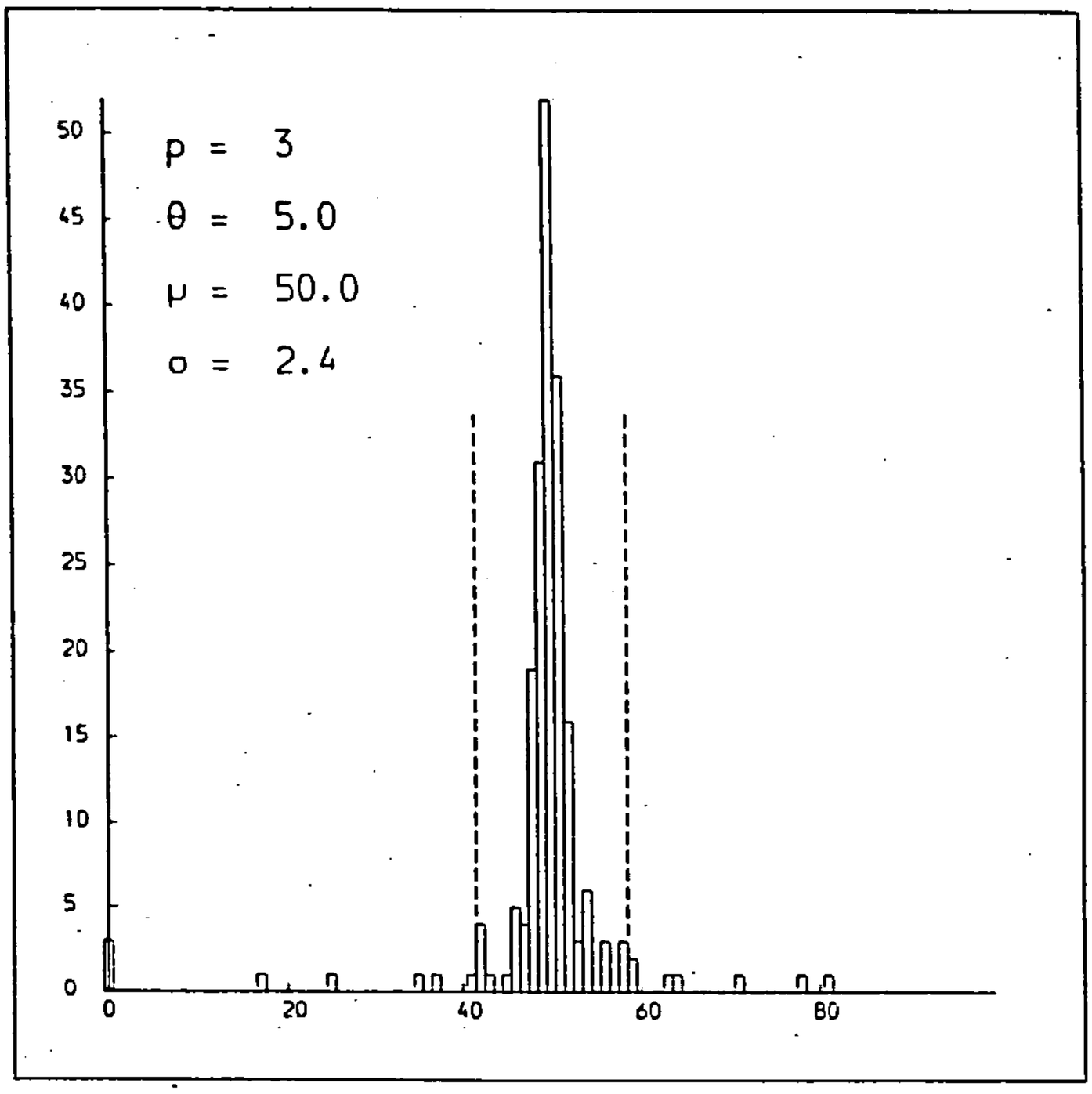
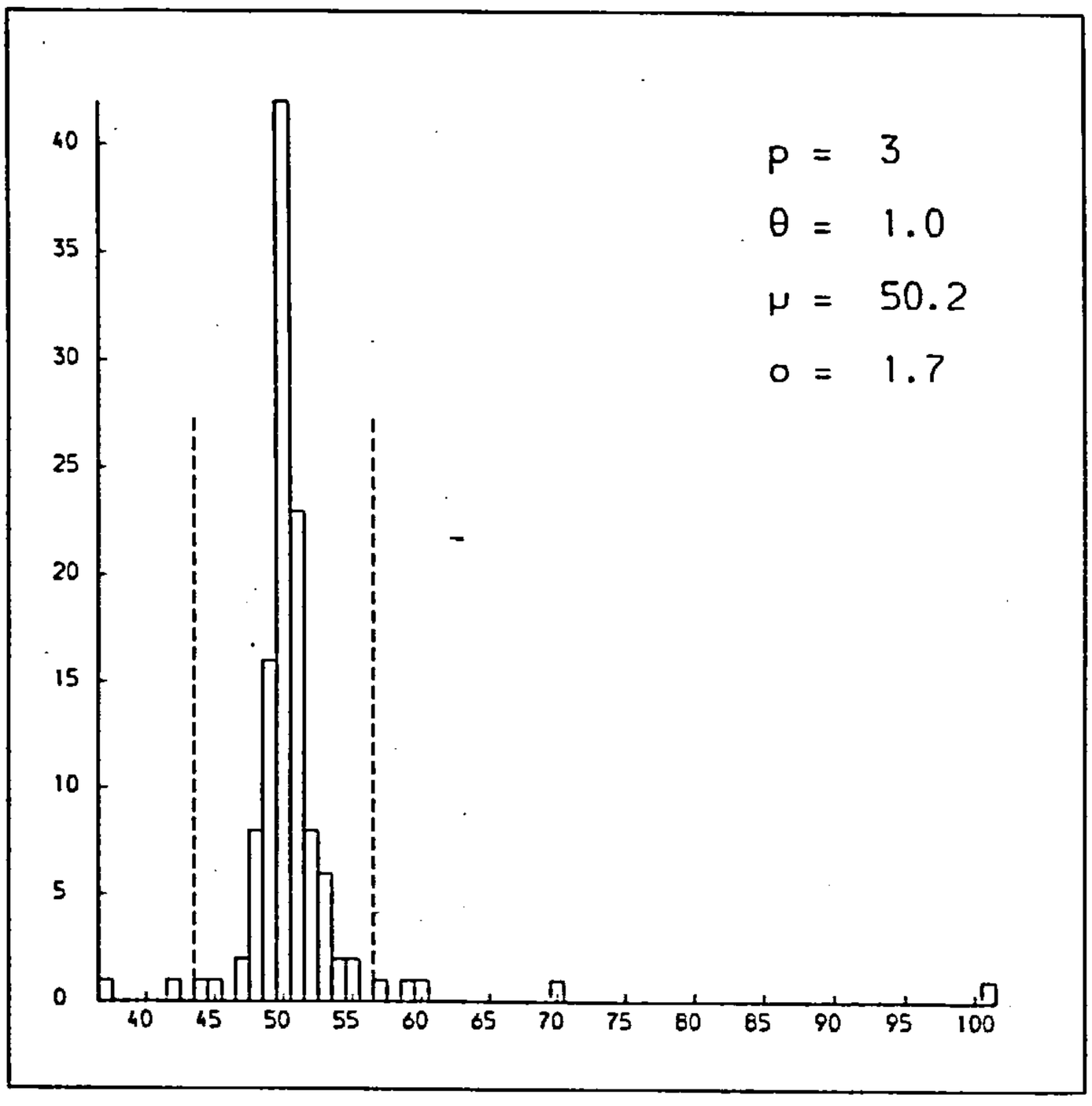


Figure 7.6; Histograms of actual syzygy counts for fictitious 3,4 and 5 planet systems. The expected mean is always 50.

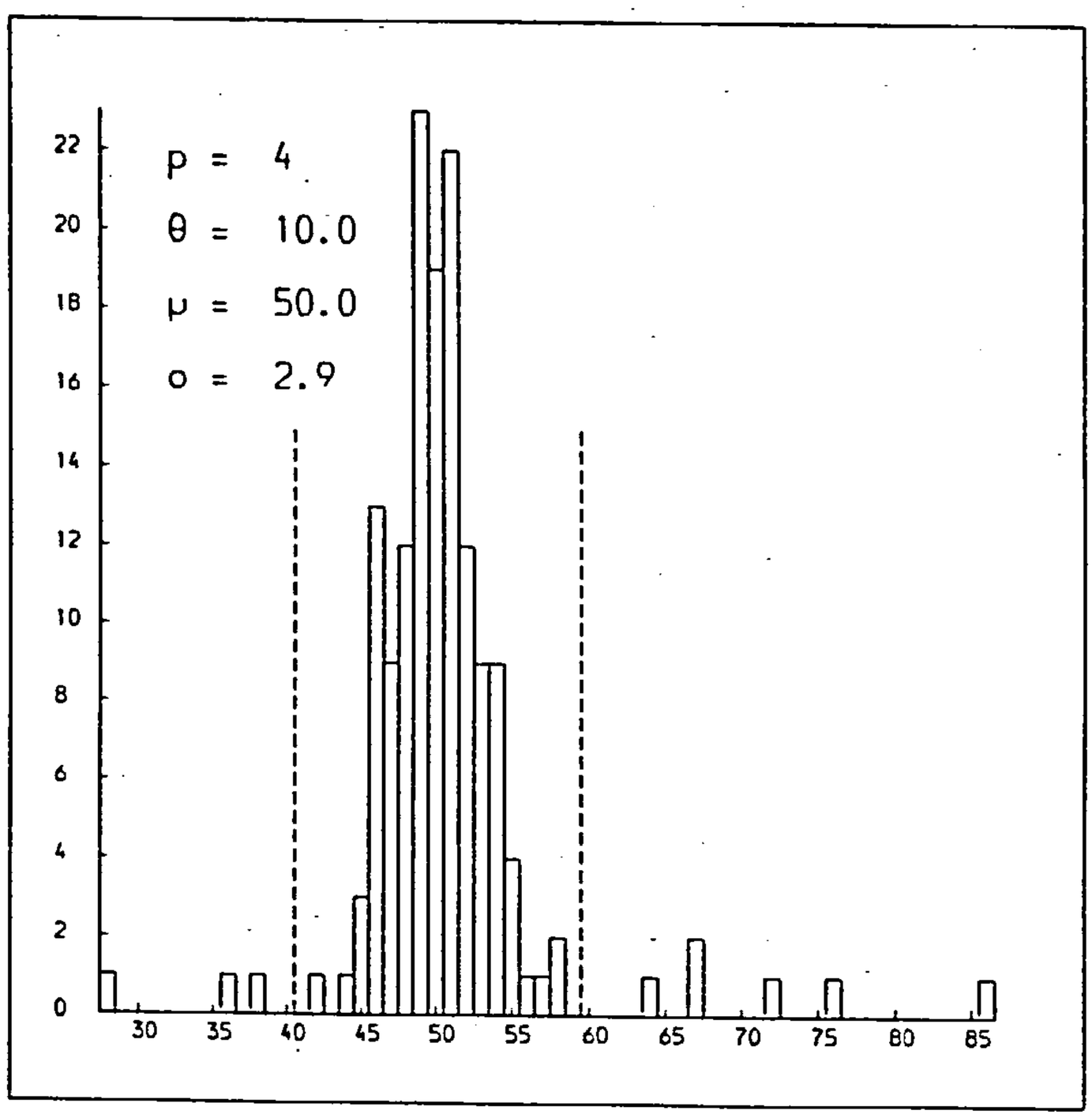
(b)


p = 3
θ = 5.0
μ = 50.0
σ = 2.4
50
45
40
35
30
25
20
15
10
5
0
0
20
40
60
80

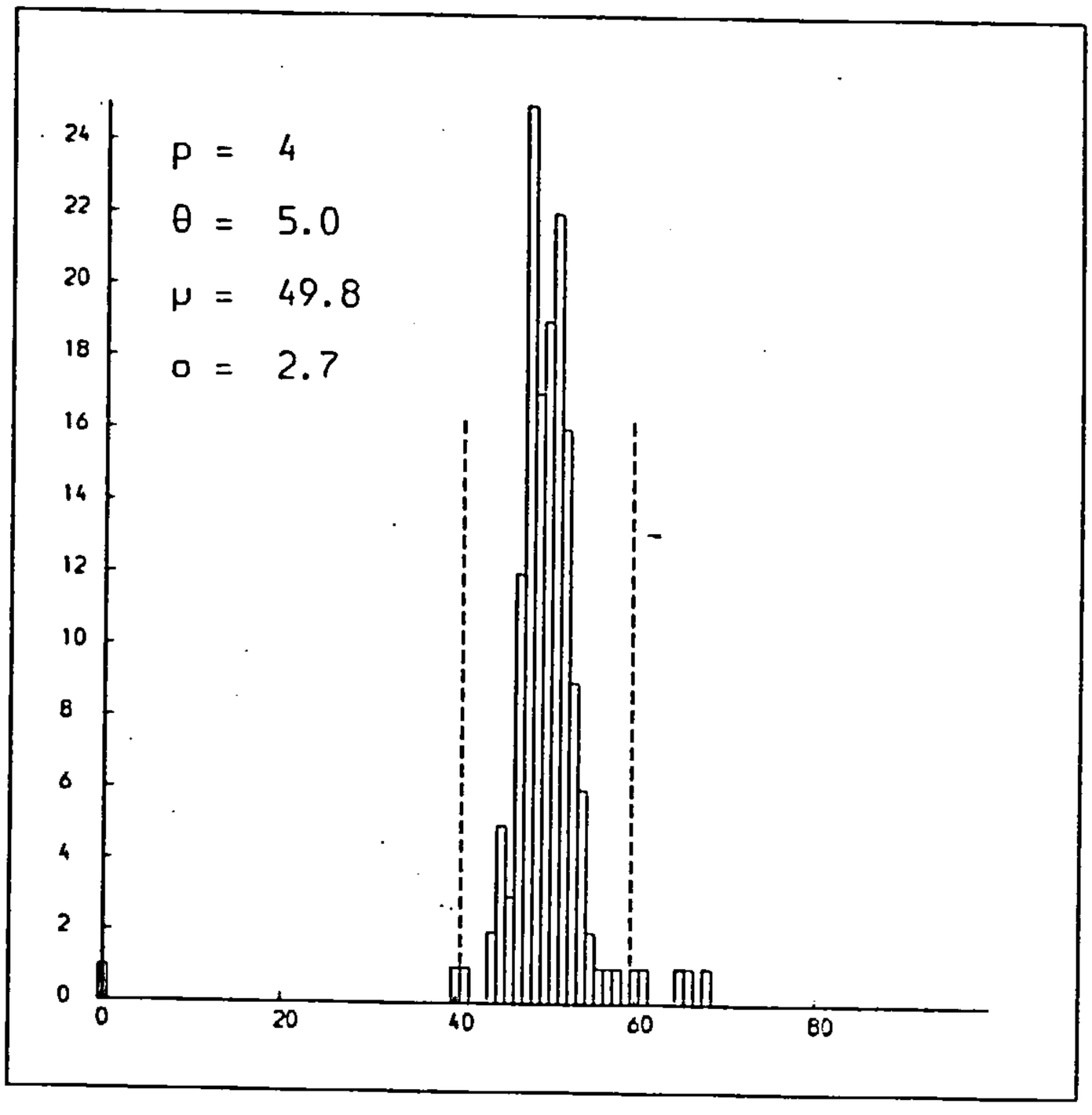
(c)


p = 3
θ = 1.0
μ = 50.2
σ = 1.7
40
35
30
25
20
15
10
5
0
40
45
50
55
60
65
70
75
80
85
90
95
100

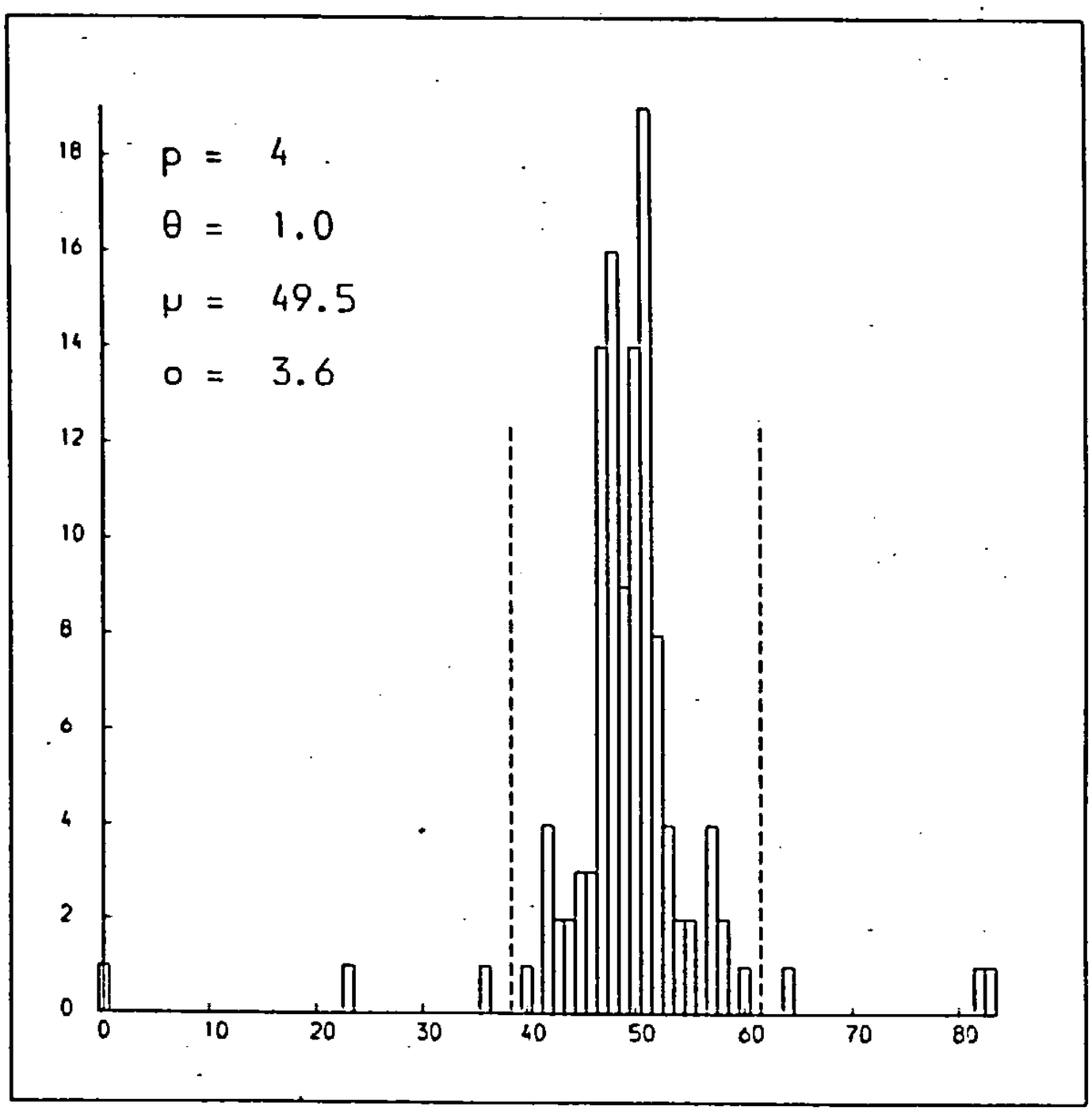
(d)


p = 4
θ = 10.0
μ = 50.0
σ = 2.9
22
20
18
16
14
12
10
8
6
4
2
0
30
35
40
45
50
55
60
65
70
75
80
85

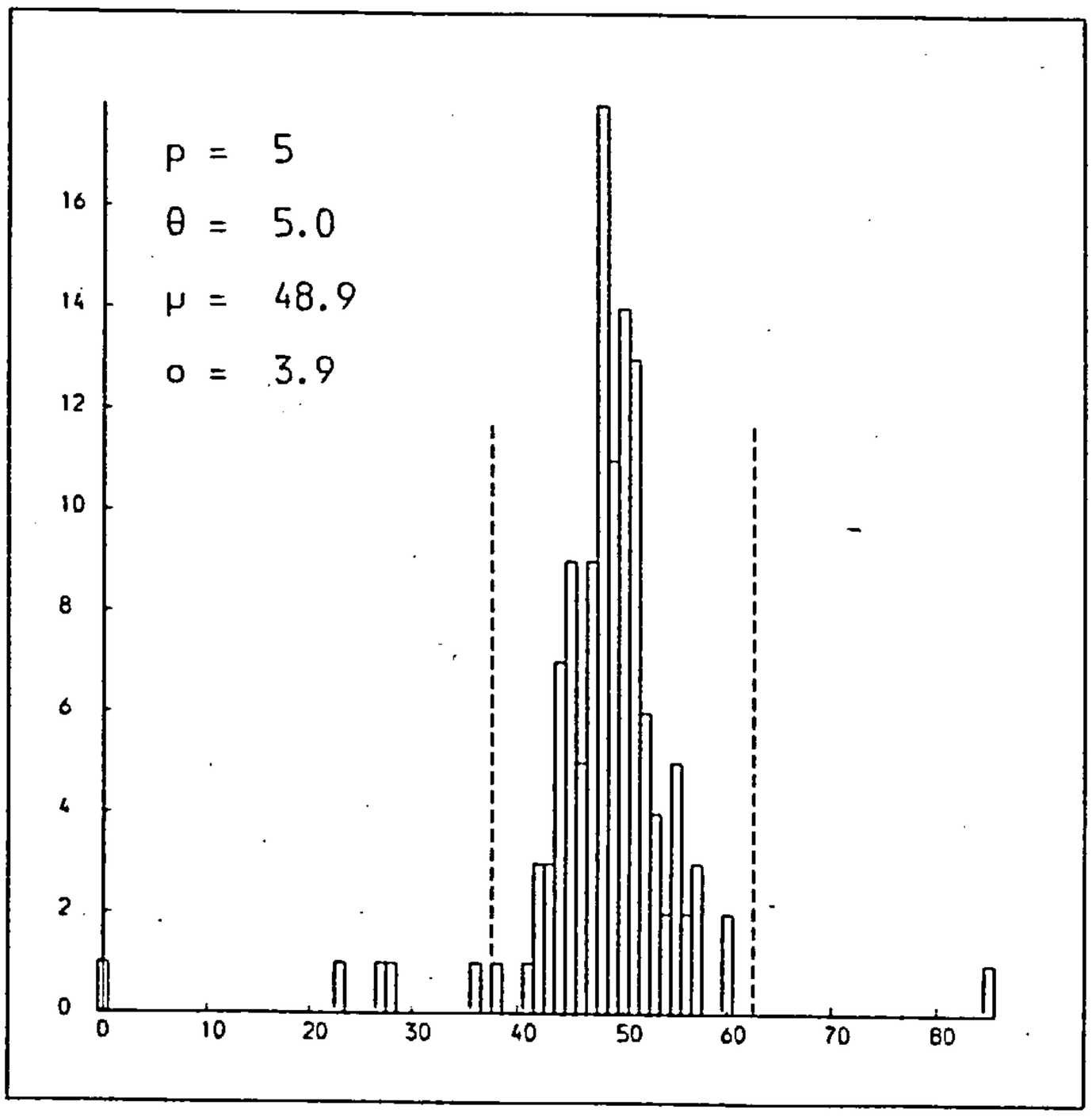
(e)


p = 4
θ = 5.0
μ = 49.8
σ = 2.7
24
22
20
18
16
14
12
10
8
6
4
2
0
0
20
40
60
80

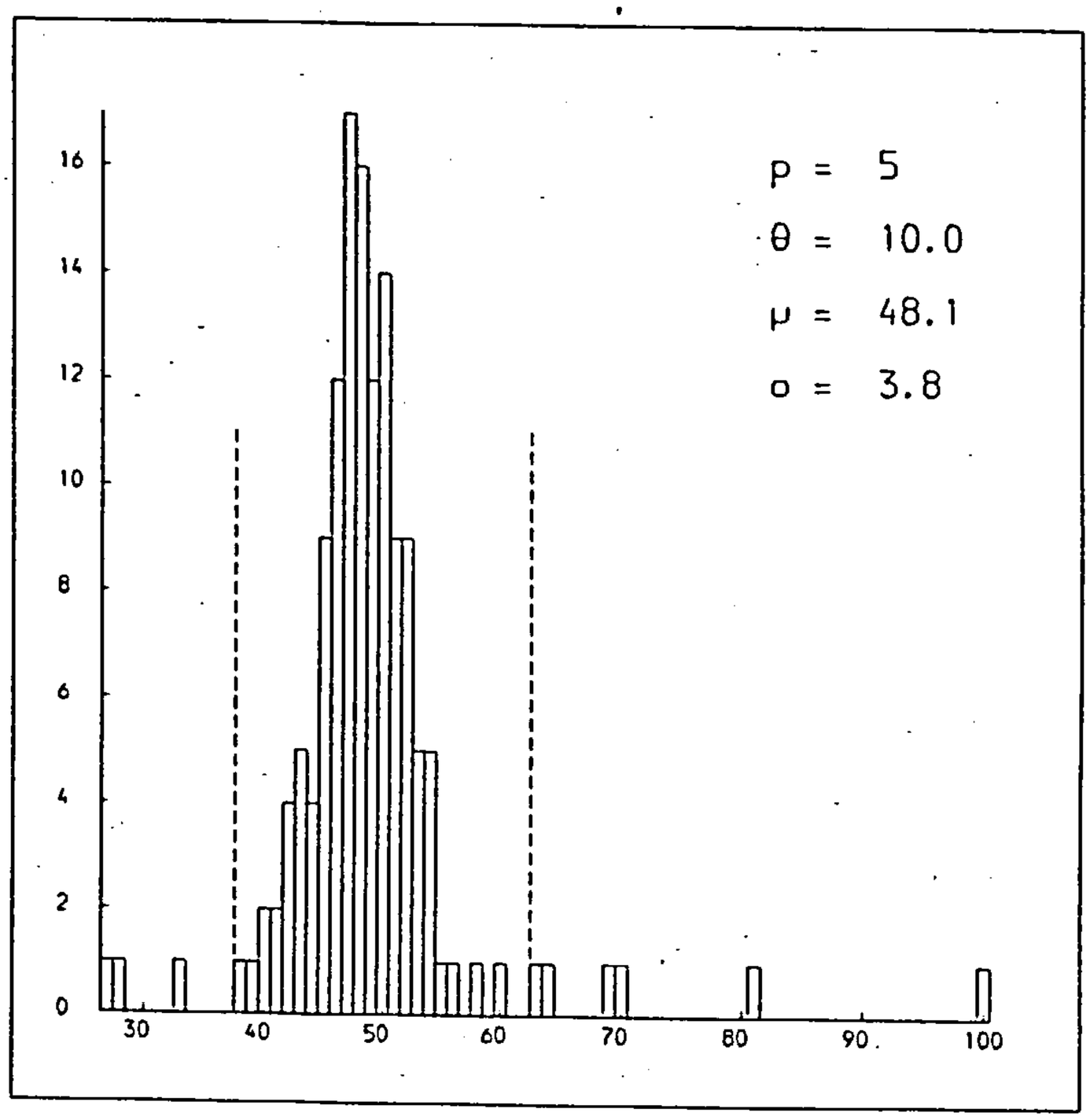
(f)


p = 4
θ = 1.0
μ = 49.5
σ = 3.6
0
2
4
6
8
10
12
14
16
18
0
10
20
30
40
50
60
70
80

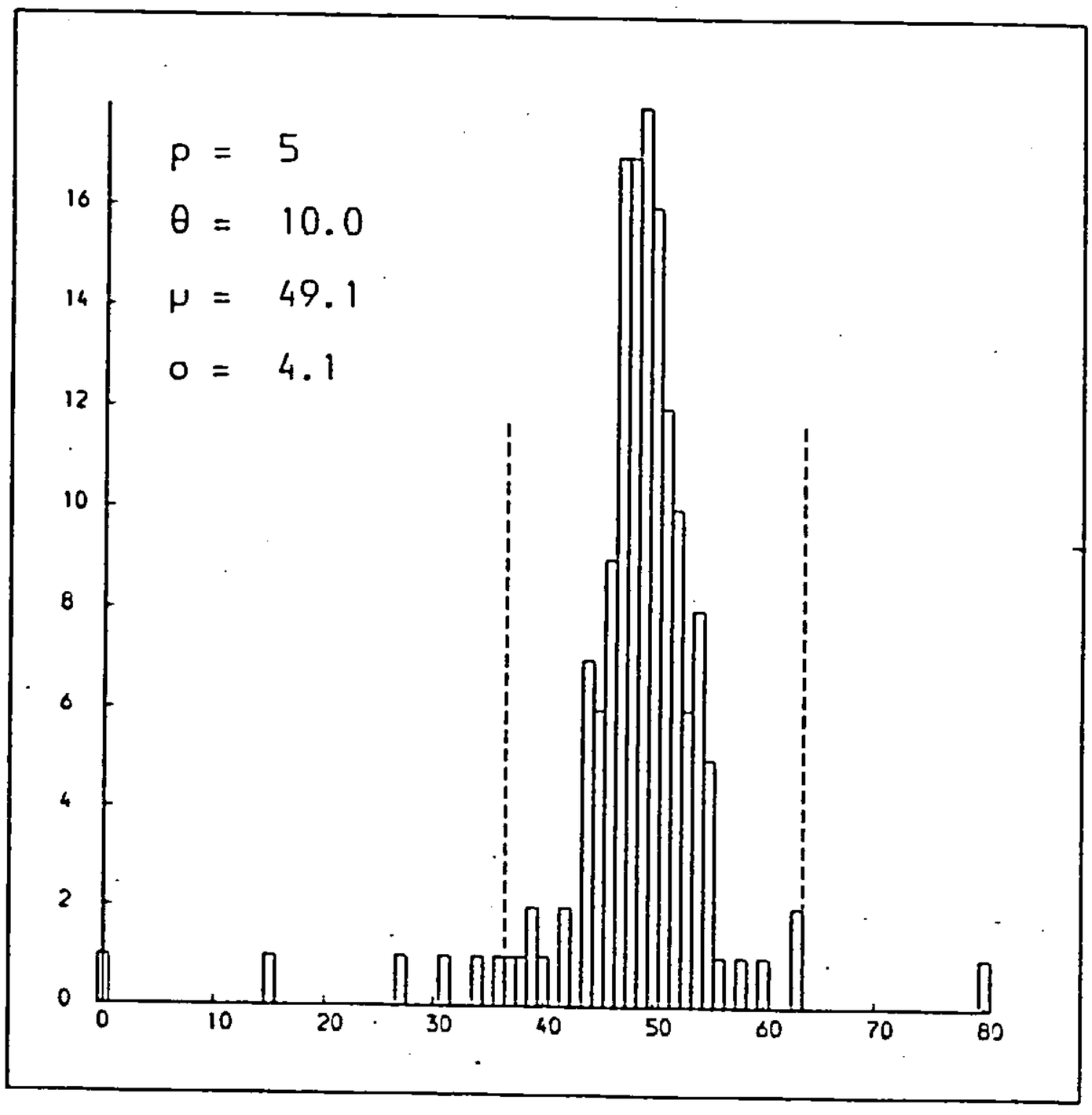
(g)


p = 5
θ = 5.0
μ = 48.9
σ = 3.9
0
2
4
6
8
10
12
14
16
0
10
20
30
40
50
60
70
80

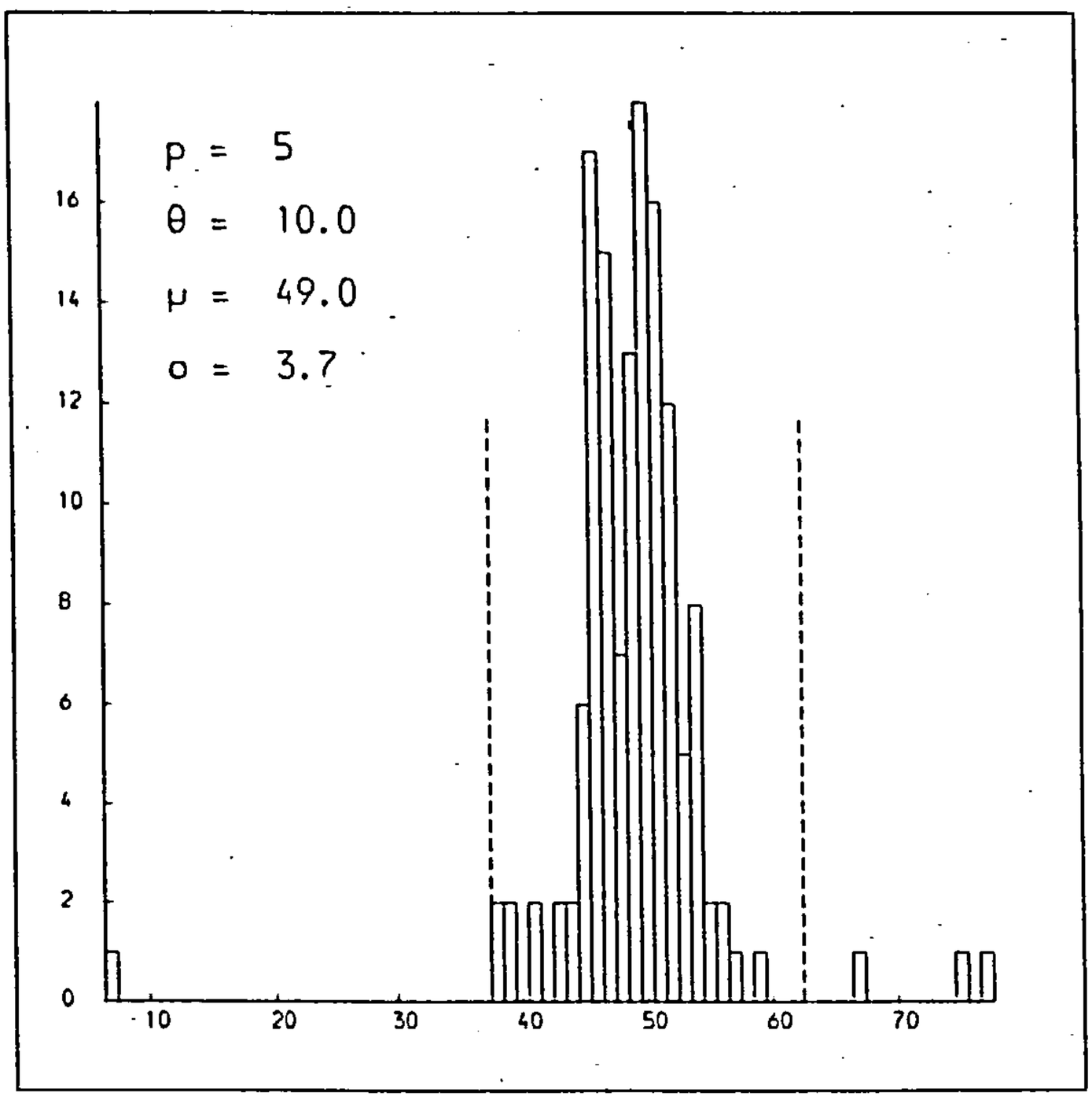
(h)


p = 5
θ = 10.0
μ = 48.1
σ = 3.8
16
14
12
10
8
6
4
2
0
30
40
50
60
70
80
90
100

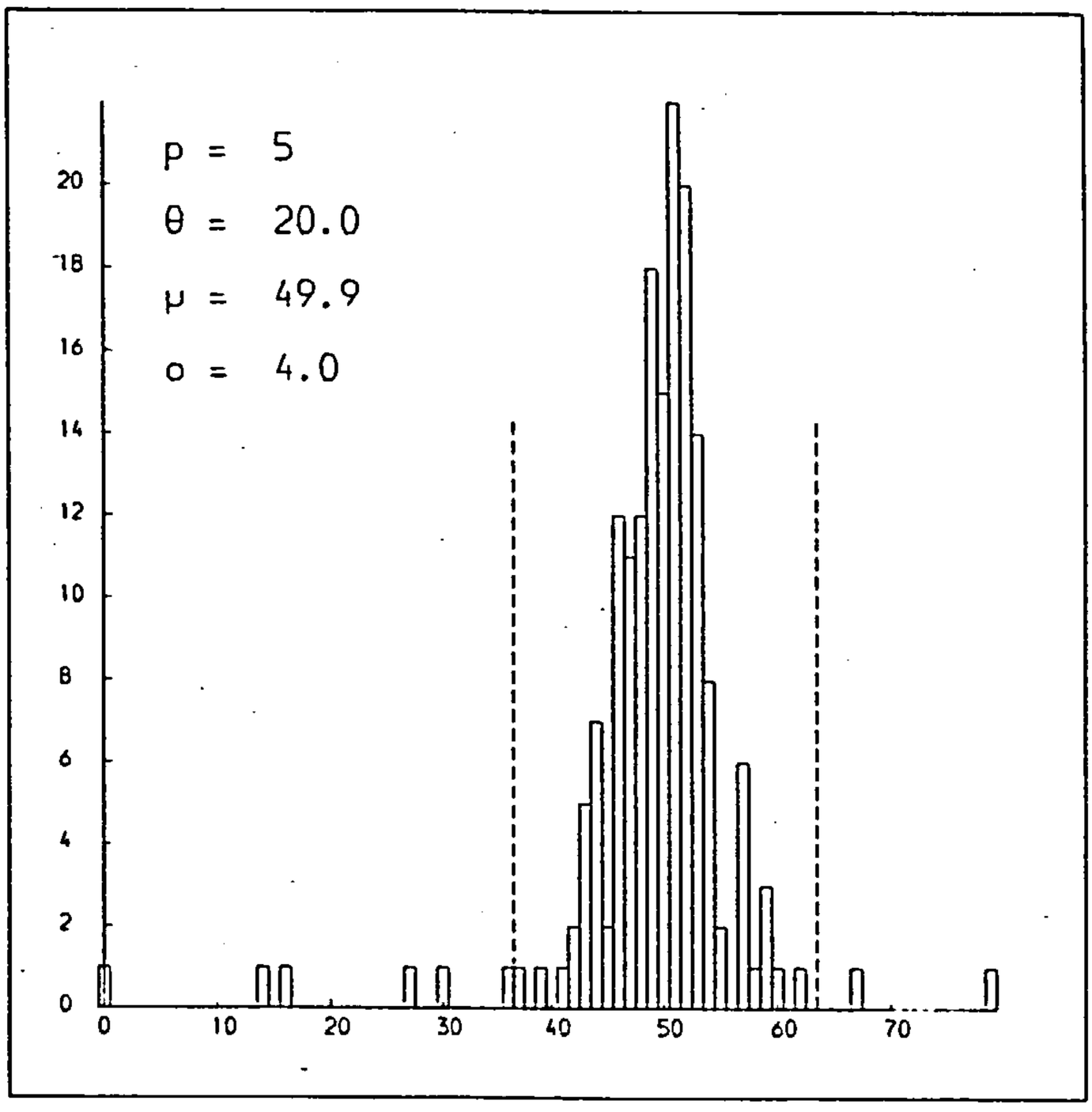
(i)


p = 5
θ = 10.0
μ = 49.1
σ = 4.1
16
14
12
10
8
6
4
2
0
0
10
20
30
40
50
60
70
80

(j)


p = 5
θ = 10.0
μ = 49.0
σ = 3.7
16
14
12
10
8
6
4
2
0
10
20
30
40
50
60
70

(k)


p = 5
θ = 20.0
μ = 49.9
σ = 4.0
20
18
16
14
12
10
8
6
4
2
0
0
10
20
30
40
50
60
70

7.3

The experiments on three and four planet systems seem to agree very well with the theory. In some cases the kurtosis is higher than it should be, indicating that the central peak is too sharp compared to the wings. Even with these reservations, the agreement seems significant. For five planet systems, problems are encountered. The data is usually normal, but the Student t-test fails. In each case, the observed mean is significantly less than 50. It is suspected that this systematic error lies in the numerical procedure, rather than in the theory. To see this, it is necessary to consider the orrery routine in more detail.

The method for detecting syzygies is to investigate the mutual differences in longitudes using two time intervals. The first coarse mesh time-step has a period corresponding to half the fastest synodic period of two of the bodies, (usually the innermost and outermost). This time-step is designed to coincide with conjunction and opposition of the relevant bodies. Around each time there will be an interval within which these bodies will be within θ^{o} of an exact syzygy. Outside this interval, there is no chance of a syzygy being detected. Within this interval we take a fine mesh of time steps and examine the mutual differences in longitudes of all other bodies to search for a syzygy.

Because we can only look for syzygies at discrete times, some occurrences may be missed, through being stepped over, thus the observed mean is less than it should be. Obviously it should be possible for the observed mean to converge to the predicted mean, by taking more and more mesh points. Indeed this seems to be the case since a mesh of 7 points gives a much poorer mean than a mesh of 11 or 51 points. When the mesh size is increased to 161 the observed mean is in very close agreement with the predicted mean. From Equation (10), the

runtime of the orrery varies with the mesh size and with Θ^{2-p}, thus in order to obtain the results for the finer meshes, it was necessary to increase the value of Θ .

In $(L_1,\ldots,L_{p-1})$ space, the shape of the region of acceptable syzygies will become more complicated as p increases. This may possibly be the reason why this discrepancy between observation and theory becomes more apparent for higher values of p. The opportunity for overstepping boundaries becomes greater. The value of Θ should have no effect, since it controls the size, not the shape, of the region.

Looking once again at Table 7.2, there seems little evidence for supposing that the standard deviation varies systematically with Θ , although much more work would be needed to confirm this. The spread in deviations does seem to widen as p increases though. If we consider the mean and standard deviation of the values of σ , for p = 3,4,5 we find

$$p = 3 \quad , \quad \sigma = 2.01 \pm 0.38$$

$$p = 4 \quad , \quad \sigma = 3.05 \pm 0.49$$

$$p = 5 \quad , \quad \sigma = 3.90 \pm 0.16$$

On applying a two-sample t-test, it is found that these differences are significant, (Hodge and Seed, 1972). This test involves calculating t where

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

where $$s = \left(\sum_{i=1}^{m} (x_i - \bar{x})^2 + \sum_{j=1}^{n} (y_j - \bar{y})^2 \right)/(m+n-2)$$

7.3

for the two samples (x_i, i=1,...,m), (y_j,j=1,...,n) and comparing this with the values given in the table of t-values in Appendix A.

Comparing the samples for p=3,4,5 we find that

$$t_{3,4} = 2.93 \quad \text{for} \quad 4 \quad \text{degrees of freedom}$$

$$t_{4,5} = 3.74 \quad \text{for} \quad 6 \quad \text{degrees of freedom}$$

$$t_{3,5} = 10.2 \quad \text{for} \quad 6 \quad \text{degrees of freedom}$$

all of which represent significant deviations at the 95% level.

Most of the evidence presented here indicates that the theory is an accurate representation of the average period of occurrence of syzygies. We have a rough measure of the distribution and deviation from the predicted values. However, if the distribution was truly normal, less than 1% of all counts should lie outside the interval ($\mu-3\sigma, \mu+3\sigma$). This is clearly not so, but it is the systems that give rise to these anomalous points that are of the most interest and which are discussed in the following sections. Any system whose behaviour deviates substantially from the theory exhibits some kind of commensurability in its mean motions. For such a system, syzygies will occur in blocks together, followed by long periods with no occurrence at all. For example, if the period of these blocks is greater than the time over which the system is being numerically investigated, then there is a chance that no syzygies will be detected. (One can see this on the histograms on Figures 7.6). By counting over longer times, the observed number of syzygies should be in closer agreement with the predicted number. This is an important point. The theory is compared with a very simple model, and will show up near commensurabilities in the mean motions over short times. Care should be taken not to over average with respect to time as this will smooth out the fluctuations.

7.3

Recall that the averaging performed in deriving Equation (10) is not only with respect to time, but also with respect to longitude as well. As examples, Figure 7.7 shows the number of syzygies observed in two three-planet systems. The crossed system is non-commensurable; the starred system is highly commensurable. Each system is started at conjunction of the inner and outer planets, but varying L_{20}, the initial longitude of the middle with respect to the outer planet. Each system is run for 20 t_3 and the graph of syzygy counts against L_{20} is plotted. It is obvious that the difference in source counts is much greater for the commensurable system. However the average over all initial configurations is very close to 20 in both cases. This indicates that the initial configuration is as important as the commensurability when it comes to detecting anomalous numbers of syzygies.

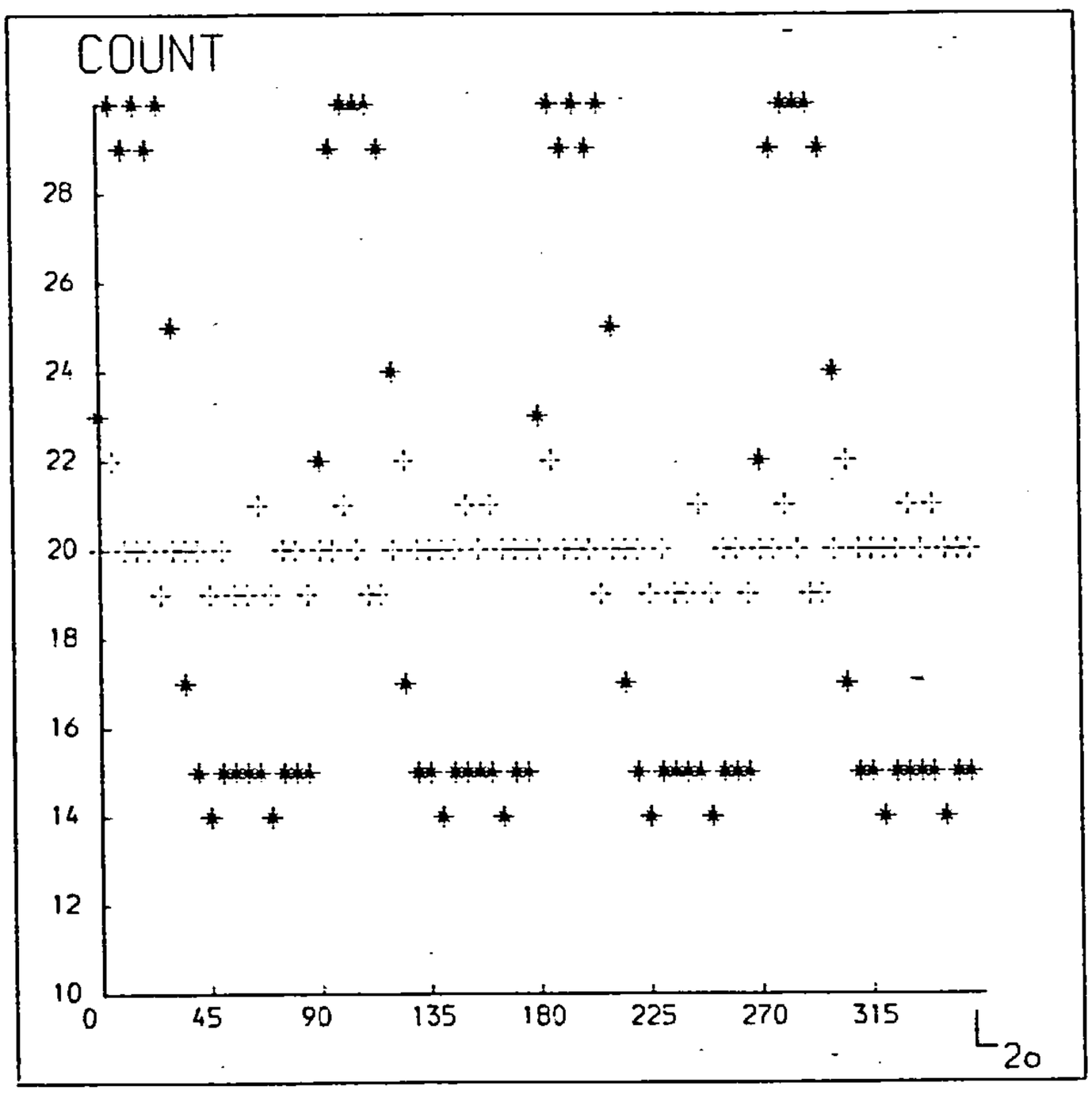


Figure 7.7 Graph of syzygy count against the initial longitude of ℓ_2-ℓ_3 for two three-planet solar systems. The "crossed" system has mean motions 0.986, 0.0831, 0.00593 (corresponding to Earth-Jupiter-Neptune). The "starred" system has mean motions 1,0.599,0.201.

7.4 Numerical Experiments for Real Systems

It is instructive to compare the behaviour of real systems with the average expected behaviour of fictitious systems. We shall restrict ourselves to only three-planet systems. Thus the equation for determining the period of syzygy occurrence is

$$t_3 = \frac{\pi^2}{\Theta} \cdot \frac{1}{2(n_1 - n_3)}$$

where n_1, n_3 are the mean motions of the inner and outer planets respectively. The numerical orrery, described in Section 7.3, is used as before, with the initial longitudes and mean motions of real four-body (three-planet) systems used as input. The results are presented for subsystems of three planets around the Sun (Table 7.3), three Jovian satellites (Table 7.4), and three Saturnian satellites (Table 7.5). There are four different types of syzygy, depending on the combination of conjunction and opposition. Remembering m_1 is the innermost planet; m_2, the middle planet; and m_3 the outermost; they are:-

Type 1: All three planets at conjunction

Type 2: m_2, m_3 at conjunction, m_1 at opposition

Type 3: m_1, m_3 at conjunction, m_2 at opposition

Type 4: m_1, m_2 at conjunction, m_3 at opposition. -

The only three planet/satellite subsystems that are considered are those with the bodies neighbouring or at most separated by one other body. It seems unlikely that any other three planet systems could exhibit a commensurability. For example, it is difficult to see how a commensurability between Jupiter, Saturn and Pluto could exist which

7.4

was not affected by either Uranus or Neptune. If a commensurability showed up in the case of Uranus-Neptune-Pluto, and Jupiter-Saturn-Pluto, it could conceivably appear in other combinations of the five outer planets. It would therefore be more constructive to consider a five planet syzygy analysis.

A commensurability may be exhibited in one of two ways. The first (Type A) is when the total number of syzygies observed deviates substantially from the expected number. The second way (Type B) is when the occurrence of two configurations are preferred over the other two (eg. Venus-Earth-Mars for $\Theta = 10^{o}$). In the first case, there is usually no doubt that a commensurability is present, albeit sometimes weakly (eg. Saturn-Neptune-Pluto for $\Theta = 10^{o}$). A little more care is needed in the second case. A subsystem with the inner moving much faster than the outer two, could conceivably orbit many times while the outer pair are close to conjunction. Thus we would see a preponderance of type 1 and 2 syzygies over types 3 and 4, or vice versa. The system of Mercury-Neptune-Pluto would exhibit such a false commensurability.

The choice of Θ is very important. In the case of a very exact commensurability like Io-Europa-Ganymede, the smaller we make Θ , the more noticeable the commensurability becomes. This subsystem finds a 1^{o} syzygy as easily as a 5^{o} syzygy, when it should be considerably harder. Because the time to find a 1^{o} syzygy is five times longer than for a 5^{o}, the actual count is five times greater.

For a less exact commensurability, the amplitude of a critical argument will be considerably greater than 1^{o}. Thus a 1^{o} syzygy search will be too stringent to detect such a commensurability. Only by increasing Θ to larger values like 10^{o}, will such commensurabilities be detected. Jupiter-Saturn-Uranus is an example of this.

Table 7.3

Syzygy counts for three-planet subsystems of the solar system. The initial longitudes and mean motions on March 31 1986 are taken from the Astronomical Almanac. The predicted number is always 20.

(a) Planet number	Name	initial longitude	daily mean motion (degrees)
1	Mercury	210.1029	4.09234
2	Venus	51.8828	1.60215
3	Earth	188.0786	0.985615
4	Mars	242.2499	0.524068
5	Jupiter	336.7161	0.0830974
6	Saturn	241.7857	0.0334005
7	Uranus	254.2854	0.0116657
8	Neptune	274.8309	0.00593309
9	Pluto	219.0029	0.00396925

(b) Planets	Θ^o	t_3(years)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
1 2 3	10	1.428	5	7	7	4	23
	5	2.855	7	4	3	6	20
	1	14.276	4	5	6	5	20
1 2 4	10	1.243	5	5	4	6	20
	5	2.486	6	5	4	5	20
	1	12.430	7	4	4	6	21
1 3 4	10	1.243	6	4	5	5	20
	5	2.486	5	5	6	5	21
	1	12.430	5	6	5	5	21
2 3 4	10	4.114	0	10	11	0	21
	5	8.228	6	6	6	6	24
	1	41.140	5	6	5	6	22
1 3 5	10	1.106	6	5	5	4	20
	5	2.212	4	6	5	5	20
	1	11.063	4	4	8	7	23
2 3 5	10	2.920	5	4	4	6	19
	5	5.840	5	4	5	5	19
	1	29.198	5	7	5	4	21

(Table 7.3 contd.)

Planets	θ^o	T_3(Years)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
2 4 5	10	2.920	3	6	6	3	18
	5	5.840	6	4	3	6	19
	1	29.198	3	6	6	2	17
3 4 5	10	4.914	5	5	4	5	19
	5	9.829	4	5	5	6	20
	1	49.143	6	6	3	4	19
2 4 6	10	2.827	6	4	6	4	20
	5	5.654	6	3	6	3	18
	1	28.273	4	6	5	4	19
3 4 6	10	4.658	4	6	4	5	19
	5	9.316	5	6	4	6	21
	1	46.578	5	5	5	5	20
3 5 6	10	4.658	4	4	5	6	19
	5	9.316	6	5	4	6	21
	1	46.578	5	4	5	6	20
4 5 6	10	9.039	6	6	4	5	21
	5	18.079	6	6	4	4	20
	1	90.393	7	6	4	4	21
3 5 7	10	4.554	4	5	5	6	20
	5	9.108	5	5	5	6	21
	1	45.539	5	5	4	6	20
4 5 7	10	8.656	5	5	6	4	20
	5	17.312	5	6	6	4	21
	1	86.558	6	5	5	5	21
4 6 7	10	8.656	6	5	5	5	21
	5	17.312	5	6	5	5	21
	1	86.558	5	6	4	5	20
5 6 7	10	62.091	8	4	4	8	24
	5	124.18	1	8	8	1	18
	1	620.91	3	6	8	4	21
4 6 8	10	8.560	6	5	5	5	21
	5	17.120	4	6	6	4	20
	1	85.601	6	7	4	4	21
5 6 8	10	57.478	4	7	6	5	22
	5	114.96	3	6	6	4	19
	1	574.78	6	4	6	5	21

(Table 7.3 contd.)

Planets	θ^o	T_3(Years)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
5 7 8	10	57.478	6	6	4	5	21
	5	114.96	6	5	5	5	21
	1	574.78	5	5	4	5	19
6 7 8	10	161.47	5	4	5	5	19
	5	322.95	4	6	4	5	19
	1	1614.7	5	5	5	5	20
5 7 9	10	56.052	4	5	4	6	19
	5	112.10	5	5	5	5	20
	1	560.52	4	5	6	5	20
6 7 9	10	150.70	5	5	6	6	22
	5	301.40	4	6	5	5	20
	1	1507.0	6	4	5	4	19
6 8 9	10	150.70	6	6	6	6	24
	5	301.40	0	10	10	0	20
	1	1507.0	7	6	6	6	25
7 8 9	10	576.27	5	4	5	4	18
	5	1152.5	5	4	6	5	20
	1	5762.7	6	5	6	5	22

Table 7.4

Syzygy counts for three of the Galilean satellites around Jupiter. The initial longitudes and mean motions on January 0.5 1900 are taken from the Connaissance Des Temps. The predicted number if always 20.

(a)

Satellite Number	Name	Initial Longitude	Daily mean motion (degrees)
1	Io	142.5999	203.4890
2	Europa	99.5508	101.3748
3	Ganymede	168.0263	50.3176
4	Callisto	234.4079	21.5711

(b)

Satellites	θ^{o}	t_3 (days)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
1 2 3	10	10.58	0	30	30	0	60
	5	21.15	0	60	60	0	120
	1	105.76	0	300	300	0	600
1 2 4	10	8.91	5	6	5	4	20
	5	17.81	6	4	6	5	21
	1	89.05	5	4	4	4	17
1 3 4	10	8.91	5	5	5	5	20
	5	17.81	0	9	10	0	19
	1	89.05	6	4	6	4	20
2 3 4	10	20.30	4	6	6	4	20
	5	40.60	5	5	4	5	19
	1	203.00	5	5	5	5	20

Table 7.5

Syzygy counts for three Saturnian satellites. The initial longitudes and mean motions on November 2.0 1960 are taken from the American Ephemeris. The predicted number is always 20.

(a)

Satellite Number	Name	Initial Longitude	Daily Mean Motion (degrees)
1	Mimas	321.04	381.999
2	Enceladus	334.78	262.7319
3	Tethys	73.175	190.6976
4	Dione	356.855	131.5349
5	Rhea	247.211	79.690
6	Titan	336.469	22.5769
7	Hyperion	179.74	16.916
8	Iapetus	56.141	4.5381

(b)

Satellite	θ^{o}	t_3(days)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
1 2 3	10	8.47	6	5	5	4	20
	5	16.94	5	4	6	4	19
	1	84.68	5	5	5	5	20
1 2 4	10	6.47	5	4	5	4	18
	5	12.94	9	0	0	9	18
	1	64.68	0	0	0	0	0
-1 3 4	10	6.47	5	5	4	5	19
	5	12.94	4	4	6	6	20
	1	64.68	4	5	5	5	19
2 3 4	10	12.35	5	5	5	5	20
	5	24.70	4	6	8	3	21
	1	123.48	6	3	4	6	19
1 3 5	10	5.36	6	5	5	4	20
	5	10.72	6	6	5	4	21
	1	53.59	4	6	5	6	21

(Table 7.5 contd.)

Satellites	θ^o	t_3(days)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
2 3 5	10	8.85	5	6	5	6	22
	5	17.70	5	4	6	5	20
	1	88.50	4	5	6	5	20
2 4 5	10	8.85	3	5	6	6	20
	5	17.70	5	5	5	6	21
	1	88.50	5	5	5	5	20
3 4 5	10	14.59	6	5	4	5	20
	5	29.19	5	5	5	5	20
	1	145.94	6	5	5	5	21
2 4 6	10	6.75	5	4	5	5	19
	5	13.49	5	6	6	5	22
	1	67.46	7	3	8	2	20
3 4 6	10	9.64	5	5	5	5	20
	5	19.27	6	4	6	4	20
	1	96.36	3	6	5	5	19
3 5 6	10	9.64	5	6	6	4	21
	5	19.27	6	5	4	5	20
	1	96.36	6	4	4	5	19
4 5 6	10	14.87	4	5	4	6	19
	5	29.74	5	5	6	4	20
	1	148.68	5	5	5	5	20
3 5 7	10	9.32	6	5	4	5	20
	5	18.64	5	5	5	5	20
	1	93.22	6	3	6	3	18
4 5 7	10	14.13	5	5	4	5	19
	5	28.27	5	5	4	5	19
	1	141.34	5	5	5	5	20
4 6 7	10	14.13	4	4	6	6	20
	5	28.27	5	4	5	5	19
	1	141.34	0	0	11	10	21
5 6 7	10	25.81	5	5	5	5	20
	5	51.61	3	6	7	4	20
	1	258.07	4	6	6	5	21
4 6 8	10	12.76	2	8	8	2	20
	5	25.51	5	5	5	5	20
	1	127.56	5	6	5	5	21

(Table 7.5 contd.)

Satellites	θ^o	t_3(days)	Syzygy Counts Type 1	Type 2	Type 3	Type 4	TOTAL
5 6 8	10	21.56	6	7	3	4	20
	5	43.11	3	4	7	8	22
	1	215.56	3	3	8	7	21
5 7 8	10	21.56	5	6	5	6	22
	5	43.11	4	5	4	6	19
	1	215.56	0	0	8	8	16
6 7 8	10	89.81	5	4	5	6	20
	5	179.61	4	5	5	5	19
	1	898.06	5	4	5	5	19

7.4

The following subsystems show evidence of commensurable behaviour:-

Venus-Earth-Mars: $n_V/n_M \sim 3$, $n_E/n_M \sim 2$.

This rather poor commensurability is still detected as a Type A for $\Theta = 5^o$ and Type B for $\Theta = 10^o$.

Jupiter-Saturn-Uranus. $n_J/n_S \sim 5/2$, $n_S/n_U \sim 3$

There is a predominance of Type 1 and 4 syzygies for $\Theta = 10^o$, but a predominance of Type 2 and 3 syzygies for $\Theta = 5^o$ and 1^o. This may indicate that the orrery quickly averages out any anomalies.

Saturn-Neptune-Pluto $n_N/n_P \sim {}^3/2$

The dominant commensurability is between Neptune and Pluto. Saturn is probably incidental. The results with Uranus-Neptune-Pluto give no evidence of a commensurability. Because the sidereal period of Uranus is longer than that of Saturn, the time per syzygy is that much longer and the short period anomalies have a chance to average out. The orrery is very inaccurate with eccentric orbits and thus any results with Pluto must be viewed with some caution. This problem will be returned to in Section 7.5.

Io-Europa-Ganymede $n_I - 3n_E + 2n_G = 0$

This famous commensurability gives a most striking example of the use of the theory and has already been discussed.

Io-Ganymede-Callisto $n_G/n_C \sim {}^7/3$

This may be a similar case to Saturn-Neptune-Pluto with Ganymede-Callisto giving the dominant commensurability.

7.4

Mimas-Enceladus-Dione $n_E/n_D \sim 2$.

There is a strong commensurability between Enceladus and Dione which is further enhanced when considered with the perisaturnium of Enceladus, rather than Mimas. (See Section 7.6). There is also some evidence for this commensurability within the Enceladus-Dione-Titan system for $\Theta = 1^o$.

Dione-Titan-Iapetus $n_T/n_I \sim 5$

There is a weak commensurability between Titan and Iapetus which only shows up for $\Theta = 10^o$.

There seems to be two false commensurabilities for Dione-Titan-Hyperion and Rhea-Hyperion-Iapetus, of the type described previously where the inner body is moving much faster than the other two. It should however be pointed out that there is a 4/3 commensurability between Titan and Hyperion but this subsystem will be examined again in Section 7.6.

7.5 Comparison of Real and Fictitious Data

For a direct comparison, the orrery was run for 1000 fictitious three-planet/satellite systems. The run time was 20 t_3 and $\Theta = 5^o$. The syzygy counts are shown in Figure 7.8. The standard deviation $\sigma = 1.5$. It should be recalled that there was no evidence that σ varied with Θ, thus we may assume this value of σ , for $\Theta = 1^o$ and 10^o. If this distribution were truly normal then 5% of the systems should lie outside the range $(20-2\sigma, 20+ 2\sigma)$, i.e. (17,23). In reality, 112 out of 1000 systems lie outside this range, i.e. 11%.

7.5

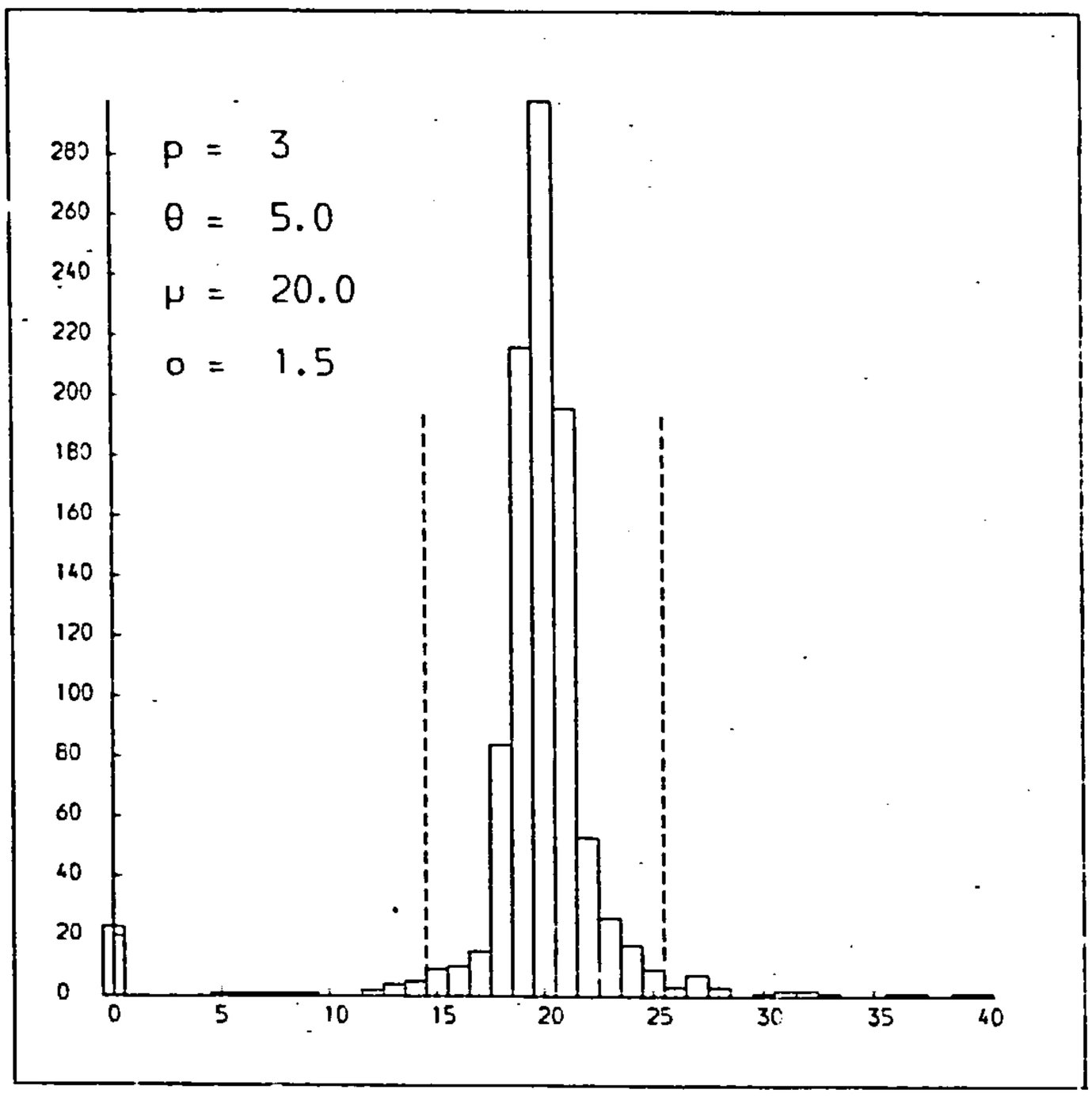


Figure 7.8 Syzygy Counts for three-planet/satellite fictitious systems. The expected mean is 20 and the dotted lines indicate the 3σ boundary. Five systems had counts greater than 40.

The real systems are significantly closer to the mean on average. Out of 48 systems, the number that lie outside the 2σ level are 3 for $\Theta = 10^{o}$ (6.25%), 2 for $\Theta = 5^{o}$ (4.2 %), 4 for $\Theta = 1^{o}$ (8.33%).

This seems to contradict the results of Roy and Ovenden (1954) where they show that there are more commensurable systems than should be expected. The reason for this is that the theory presented here does not take into account the fact than in the solar system, the orbits are well spaced. None of the relative mean motions N_i are close to zero so there is a bias against such systems as will go for long periods without encountering a syzygy. It is the fictitious systems with slow relative mean

7.5

motions that give the zero counts in Figure 7.8, (23 in all). If they were run for longer, we would observe a sudden "burst" of syzygies grouped together. The outer bodies with slow relative motion finally come close enough, and stay close for a long time. The inner body may be moving very quickly by comparison, and performing many orbits during the one conjunction of the outer bodies. Thus after the famine comes a feast of syzygies occurring every sidereal period of the inner body!

It would therefore be quite wrong to expect as large a spread of syzygy counts within the real solar system as are displayed by the fictitious systems. It is noticeable however that the orrery as used here is far less effective in dealing with eccentric orbits. Possible remedies are discussed in Sections 7.6 and 7.7.

7.6 The Search for Critical Arguments and Mirror Configurations

Everything said so far has been related to occurrences of syzygies in unperturbed, coplanar, circular, planetary orbits. If the orbits are given small eccentricities and are perturbed by the other bodies in the system, then the situation is considerably more complicated, since the osculating mean motions are now changing in a periodic manner. Our assumption of constant angular velocity is no longer valid. This means that the straight line describing the evolution of the system (Equation (4)) now becomes a helix with all the difficulties in calculating cross-sections that are implied. Providing the system is stable within the observed time, however, the average values of the mean motions are likely to remain fairly constant. This theory can

7.6

therefore still be applied, for small eccentricities. These eccentricities do however raise many new and interesting possibilities.

Some subsystems exhibit critical arguments where the conjunction line of two bodies oscillates about the moving apse line of one of the bodies. As examples,

(a) Neptune-Pluto-Pluto's apse, has critical argument

$$\Theta = 3\ell_P - 2\ell_N - \tilde{\omega}_P.$$

It is found that Θ librates about 180° with an amplitude of 76° and a period of 19,670 years.

(b) Titan-Hyperion-Hyperion's apse gives

$$\Theta = 4\ell_H - 3\ell_T - \tilde{\omega}_H \sim 180^\circ$$

with an amplitude of 36°.

As will be discussed in Section 7.7, these resonances may be maintained by a dynamical mechanism which is not included in the simple numerical orrery described here. However, over relatively short times, such resonances will have a marked effect on the number of syzygies detected and perhaps lead to the discovery of new critical arguments.

To investigate resonant systems, the apses must be considered and apsidal as well as planetary syzygies looked for, the pericentres being considered as separate "bodies". Thus the argument and mean motion of pericentre can be used in an analogous manner to the longitude and mean motion of a real body.

For example, it is possible to examine the 4:3 commensurability of Titan and Hyperion, by running the orrery for

7.6

$$L_1 = \ell_T - \tilde{\omega}_H \quad , \quad N_1 = n_T - \dot{\tilde{\omega}}_H \ ,$$

$$L_2 = \ell_H - \tilde{\omega}_H \quad , \quad N_2 = n_H - \dot{\tilde{\omega}}_H \ ,$$

as found in the Explanatory Supplement to the Ephemeris. From the epoch 1900.0, in the time predicted for 20 syzygies within $\Theta = 10^o$ (t_3 = 72 days), there are in fact 45.

An even more striking example is given by Enceladus-Dione-Enceladus apse. The initial conditions on April 0.5 1889 are taken from the Connaisance Des Temps. In the time predicted for 20 syzygies within $\Theta = 2^o$ (t_3 = 31 days), there are 450.

For many orbits with very low eccentricities, the rate of precession of the apses are very poorly determined. In any case, it is debatable if the apse of such a system plays a significant role in a critical argument.

For high eccentricies ($e \gtrsim 0.1$), this method for finding critical arguments becomes suspect due to the variations in angular velocity of the bodies involved, and a more accurate numerical procedure may be necessary, (see Section 7.7).

This use of apsidal syzygies is helpful in considering near mirror configurations, which are described by Roy and Ovenden (1955) as being configurations when all the mutual radius vectors are perpendicular to all mutual velocity vectors within the system. If such a configuration occurs then the behaviour of the system after that time is a mirror image of its behaviour before that time. If two exact mirror configurations occur in a system's lifetime, it is periodic. For a coplanar system of elliptic orbits the condition for a mirror configuration is that all the radius vectors and apse line vectors must be parallel. Thus to search for mirror configurations, it is necessary

7.6

to look for syzygies of bodies and apses.

Table 7.6 gives the average period of occurrence of bodily syzygies and mirror configurations in the solar system. By inverting Equation (10), it is found that in the age of the Solar System ($\sim 4.5 \times 10^9$ years), the closest syzygy we can expect is for $\Theta = 4^o$, and the closest syzygy of bodies and apses should be approximately 25^o. In other words, there should be little chance of a "good" mirror configuration. Note that although p = 3 in Equation (10) when deriving the period of syzygy occurrence for Jupiter-Saturn-Uranus, p = 6 is used to derive the period of occurrence of mirror configurations. Due to the low eccentricity of Venus and Neptune, their apses are neglected (ie. p = 16 to calculate the period of occurrence of mirror configurations for all nine planets). Pluto causes problems. Not only are the eccentricity and inclination of its orbit high, thus violating the assumptions of coplanar, circular orbits, but the rate of precession of the apse is not well determined. Thus any results incorporating Pluto must be viewed with some caution.

	Θ^o	syzygy period (years)	mirror configuration period (years)
JSU	5	124	1.57×10^6
JSUN	5	2520	4.55×10^7
JSUNP	5	61900	4.18×10^{10}
Solar System	10	6.40×10^6	$\sim 10^{15}$
Solar System	5	8.19×10^8	$\sim 10^{19}$
Solar System	2	5.00×10^{11}	$\sim 10^{25}$
Solar System	1	6.40×10^{13}	$\sim 10^{29}$

J=Jupiter, S=Saturn, U=Uranus, N=Neptune, P=Pluto

Table 7.6 Average period of occurrence of planetary syzygies and mirror configurations for the nine planets of the Solar System and subsystems of the five outer planets.

7.7 Discussion

This chapter has been concerned with developing methods for detecting syzygies within the Solar System. Equation (10) provides a gauge of the frequency of occurrence of syzygies, which has in turn led to methods for discovering commensurabilities, critical arguments and mirror configurations, where the orbits have small eccentricities.

If however the eccentricities become large, the angular velocities of the bodies are no longer approximately constant and the theory becomes invalid. The simple orrery described in Section 7.3 is no longer sufficient. Given a long enough time, this routine will average out all inexact commensurabilities which is why the orrery was never run for longer than 20 t_3 when examining real systems. Thus this theory is valid only for low eccentricities, and its application to the discovery of commensurabilities is valid only for short times.

In addition there is the possibility that a dynamical mechanism may be encouraging a commensurable system. Goldreich (1965) has suggested that tidal forces may stabilize commensurable systems, thereby allowing frequent occurrences of syzygies for all time. In this case no amount of averaging with respect to time will allow the observed mean to tend to the expected mean. It is obvious that the simple numerical procedure indicated here could not detect such a "locking mechanism". To investigate such a dynamical effect, we would require a much more complicated numerical integration procedure.

Over the years, there have been many numerical schemes for integrating the equations of motion of the Solar System. Some of these have been discussed in Section 1.6. Recently, a Longterm Gravitational Study of the Outer Planets (LONGSTOP) project has been

7.7

proposed (Roy, 1983). This project incorporates some dynamical effects missing from the classical six-body problem, and will examine the five outer planets over 10^8 years, (5×10^7 back and forward from the present time). The inner planets are simulated by a ring around the Sun and relativistic effects are considered.

In order to overcome the difficulties in detecting syzygies described here, it will be necessary to use such a numerical procedure. Data on the orbits will appear relatively infrequently due to problems of storing so much information. A large step length is also preferable from the point of view of computational speed. However, the simple orrery may be used to interpolate between the points, where an assumption of constant angular velocities may be sufficiently accurate.

Obviously the simple resonances given here as examples can be detected by an examination of the mean motions and longitudes of the relevant bodies and apses. If we consider systems with more than three bodies it becomes difficult to find more subtle resonances.

In this way, an exhaustive search for commensurabilities, critical arguments and mirror configurations may be carried out for very large numbers of objects, be they bodies or apses. This would be a major project in itself and is outside the scope of the present work.

CHAPTER 8

UNANSWERED QUESTIONS

8.1 Refinements in the Analysis of Results from the Three-Body Numerical Experiments

8.2 Comparison of Results with General Perturbation Theory

8.3 Cross-Overs, Escapes and Close Encounters

8.4 Further Three-Body Numerical Experiments

8.5 The Four and More Body Problem

8.6 The Search for Syzygies

8.7 Final Remarks

It is inevitable, in a study such as this, that there remain many unanswered questions. Some have arisen during the course of this work. Some have existed for many years previously. This final chapter is concerned with looking at some of the more important questions and suggests methods for investigating the answers.

8.1 Refinements in the Analysis of Results from the Three-Body Numerical Experiments.

In Chapters 4 and 5, results from several hundred numerical integration experiments were given for coplanar, initially circular three-body systems. In particular, graphs were given of stability lifetime N_s against the initial ratio of semi-major axes α, grouped according to the values of ε^{23} and ε_{32}, the empirical stability parameters. Analytic curves were fitted to the (α, N_s) data to allow subsequent predictions of stability lifetime, given ε^{23}, ε_{32}, α.

Due to the smooth nature of the data, a thorough investigation of the unstable retrograde systems was possible within the chosen ranges of initial conditions. This allowed accurate values of α_o, the limit of hierarchical stability to be determined. A similar systematic search was not performed with the direct systems for two reasons. Firstly, much of the data was taken from Walker and Roy, Paper III and secondly, the range of initial conditions for unstable direct systems is much greater.

A more thorough numerical examination of the direct systems is highly desirable, for many reasons. It was pointed out in Section 5.4, that it was often difficult to say whether the observed asymptote at α_o was real or whether it was a strong commensurability and more unstable systems existed for lower values of α. By investigating more

systems near the suspected value of α_o, many of these questions may be resolved. It may also be prudent to increase the accuracy of the numerical integration method, in order to study the systems for longer and avoid the worry of numerically induced commensurabilities (Section 5.4). In any case, with more data, the precision of the lifetime predictions should also increase (Section 5.5).

The curve that was fitted to the (α, N_s) data was of the form

$$N_s = f(\alpha) = \exp\left[\beta\left(\frac{1-\alpha}{\alpha-\alpha_o}\right)^{\gamma}\right]-1 \quad (\alpha_o < \alpha \leq 1) \qquad (1)$$

The three fitted parameters are α_o, β, γ (α_o being the boundary value of α for hierarchical stability). These parameters are all functions of ε^{23} and ε_{32}. For both direct and retrograde systems, α_o was seen to rise monotonically as ε^{23}, $\varepsilon_{32} \to 0$. Unfortunately, no such trends were observed for β or γ. There could be two reasons for this. The first (and more likely) is that the errors in estimating β and γ are still large and any systematic trend is swamped by the statistical error. The second possibility is that $\beta(\varepsilon^{23}, \varepsilon_{32})$ and $\gamma(\varepsilon^{23}, \varepsilon_{32})$ are genuinely not monotonic functions. This question may possibly be resolved if more data aremade available. Another alternative may be to choose a different form for the curve. It is seen from Equation (1), that $f(1) = 0$ at present. It was pointed out in Chapter 4 that there exists a value of α, denoted by α_x, at which the orbits of the inner and outer binary cross.

It can be shown that

$$\alpha_x = \frac{1}{1-\mu} = 1 + \varepsilon^{23} \qquad (2)$$

implying $\alpha_x > 1$. In reality therefore $N_s = 0$ at $\alpha = \alpha_x$, and $N_s > 0$

at $\alpha = 1$. On studying Figures 4.1, it is clear that $f(\alpha)$ while adequate is not an ideal form for the best fit curve to the (α, N_s) data. There is a systematic error which cannot be overcome by changing the values of α_o, β and γ alone. A more accurate form may therefore be

$$f_1(\alpha) = \beta \frac{(\alpha_x - \alpha)^\gamma}{(\alpha - \alpha_o)^\delta} \qquad (3)$$

or

$$f_2(\alpha) = \exp\left[\beta \frac{(\alpha_x - \alpha)^\gamma}{(\alpha - \alpha_o)^\delta}\right] - 1 \qquad (4)$$

where the parameters $\alpha_o, \alpha_x, \beta, \gamma, \delta$ are functions of ϵ^{23} and ϵ_{32}. α_x is a known function of ϵ^{23} alone (Equation (2)) so only $\alpha_o, \beta, \gamma, \delta$ need to be determined by statistical methods. There are now four parameters instead of three, making the statistical analysis more complicated. It is unlikely that one of the forms in Equations (3) and (4) would produce significantly better results than the other. Equation (4) is the more complicated but would produce a greater spread of points over the x-range during the normalisation procedure described in Section 5.5.

It may be of interest to group the (α, N_s) data according to equal values of the normalised masses μ, μ_3 rather than ϵ^{23}, ϵ_{32}. The curve fitting procedures are general enough that they could still be applied and predictions of stability lifetimes could be obtained. Note that the parameters α_o, β, γ would be functions of μ, μ_3 rather than ϵ^{23}, ϵ_{32}. The advantage of grouping in this way is to make comparisons with real systems easier (see Section 6.4). A real system may be characterised by the values of μ, μ_3, α (assuming it is coplanar and initially circular). It would be useful to obtain $\alpha_o(\mu, \mu_3)$ to

say at what value of α the system becomes unstable. If α_o is given in terms of ε^{23} and ε_{32} , this value becomes less clear, since ε^{23} and ε_{32} vary with α. A disadvantage of the (μ,μ_3) grouping is that the curves are likely to be less sharp and the values of α_o correspondingly less well defined. There is liable to be significantly greater ranges of α where the systems are unstable but nevertheless survive for many synodic periods. This was the reason why Walker and Roy originally chose to group according to ε^{23} and ε_{32}. Another disadvantage is that the range of possible values of μ,μ_3 is greater than the range of ε^{23} , ε_{32} ; $(0 < \mu \leq \frac{1}{2}$, $0 < \mu_3 < \infty)$ compared with $(0 < (\varepsilon^{23}$, $\varepsilon_{32}) < 1)$. This implies that many more (μ,μ_3) groupings must be studied.

8.2 Comparison of Results with General Perturbation Theory

Section 4.5 considered a few retrograde systems in detail. By numerical integration, graphs were obtained showing the variation with time of the semi-major axes and eccentricities of the inner and outer binary orbits. Two curves were superimposed on the graphs, showing the behaviour of the elements at every even conjunction and every odd conjunction, respectively. From these graphs, the differences in behaviour between stable and unstable systems were clearly seen. Other effects were also apparent. The odd conjunction curves were seen to oscillate with opposite phase to the even conjunction curves. The amplitude, period and initial phase of these oscillations depended on the sizes of ε^{23} and ε_{32} and also whether ε^{23} was less than, equal to, or greater than ε_{32} . There was some evidence to suggest that the period may be related to the precession of the apses. Another interesting difference can be seen between the behaviour of the eccentricities

e_2, e_3 for the inner and outer binaries, respectively. Between conjunctions, e_3 has four turning points, while e_2 has only two.

On examining the retrograde systems in Figure 4.1, we observe that N_s rises as α decreases. For $\varepsilon^{23} \gtrsim \varepsilon_{32}$, the rise is far more dramatic than for $\varepsilon^{23} < \varepsilon_{32}$. In most graphs, N_s rises monotonically, the only exceptions being for $\varepsilon^{23} = 10^{-2}$, $\varepsilon_{32} \lesssim 10^{-4}$ when an anomalous peak in N_s appears at high α .

We can only guess at the physical reasons behind many of these effects until a more detailed study is performed. The first objective of this thesis has been to obtain enough data from numerical integration experiments to perform reliable statistical curve fitting and hence be able to predict stability lifetimes for three-body systems. A more detailed analysis of individual systems would have slown down the process at the time, but would now be of considerable interest.

Such a study could involve the results from numerical integration compared to a general perturbation theory, (see Section 1.5). Such an approach could yield information concerning the periodicities observed in the retrograde systems. More graphs of the orbital elements including the arguments of pericentre should be obtained by numerical integration of the equations of motion or by numerical integration of the Lagrange planetary equations (Section 1.6). An analytical perturbation theory may provide information concerning the periodicities observed in retrograde systems. Similar studies could be applied to the direct systems in order to understand more clearly which commensurabilities are important in enhancing stability (Section 5.4).

8.3 Cross-Overs, Escapes and Close Encounters.

The definition of hierarchical stability requires three separate

conditions to be satisfied: no body must escape; no orbits should ever cross; there must be no secular changes in semi-major axis, eccentricity or inclination of the orbits ("close encounter").

The retrograde three-body systems seem to approach instability in a consistent manner (Chapter 4) with secular changes in the elements leading to a crossover of orbits. The behaviour of the direct systems, on the other hand, shows greater variation. This makes it difficult to classify the manner of instability and leads to subjectivity in determining the stability lifetime of an individual system (Section 5.2).

Chapter 3 reviewed at length the analytical Hill-type stability criterion applied to the general three-body problem. By this method, sufficient conditions can be derived which prevent a cross-over of orbits. When this happens, the topology of the zero-velocity curves prevent the third mass from approaching the other two. Marchal (1985) describes this state as *isolated*. When the zero-velocity curves allow the third mass to approach either of the other two, an *interplay* may take place.

The treatment by Marchal and Saari allows less stringent conditions to be derived. Using the notation of Section 3.2, the Sundman Inequality is

$$\frac{\sigma}{\nu} \geq j \equiv -\frac{p}{2\sigma} + \frac{\sigma}{2a} + \frac{\sigma\,\sigma'^2}{2GM} \quad .$$

Marchal (1985) points out that whenever the Sundman function $j \geq \sigma/\nu(L_1)$, the third body m_3 is isolated. (If m_3 is the smallest mass, it suffices that $j \geq \sigma/\nu(L_2)$). The minimum value of j is $\sqrt{p/a}$. If $\sqrt{p/a} \geq \sigma/\nu(L_1)$, (or $\sigma/\nu(L_2)$), the isolation is permanent and the system shows the hierarchical preservation described in Chapter 3*.

---

*With so many terms being used, it is worth recalling that "no cross-over", "hierarchical preservation" and "permanent isolation" are synonomous.

If $\sqrt{p/a} < \sigma/\nu(L_1)$(or $\sigma/\nu(L_2)$), the isolation may only be temporary and interplay can occur.

From the results of this thesis, it seems there exist systems where the bodies are only temporarily isolated but which never actually interplay in such a way as to cross orbits. It would be of interest to monitor the value of j for systems with initial α in the range (α_c,α_o) to examine why m_3 never approaches m_1 or m_2. As an example, if j is large and $\sigma' > 0$, j is non-decreasing so long as σ is non-decreasing. In this case a temporarily isolated body could remain isolated for a considerable length of time.

Marchal (1985) also gives conditions for escape, stating "A$_h$ isolated body that, at any time, has an escape velocity arrives from infinity and/or will escape to infinity."
He then proceeds to derive bounds on the escape velocities. It would be interesting to compare these conditions with the numerical experiments performed here. Very few escapes have been detected in these experiments but this is due to the system exhibiting either a cross-over or close-encounter first. If the numerical integration was allowed to continue past this point, more escapes may be apparent. Marchal shows that the smallest mass is mostly likely to escape the system. In practice we are unable to find a system where one of the larger masses escapes.

These conditions for isolation and escape are extremely relevant to the work presented here. In Section 3.3, α_c was calculated for various pairs of ε^{23} and ε_{32}. m_3 is permanently isolated for any system with initial α in the range $(0,\alpha_c)$, but this does not rule out the possibility that m_3 could escape. For most pairs of ε^{23}, ε_{32} an empirical limit on hierarchical stability α_o is found which

exceeds α_c. This implies that the probability of m_3 escaping becomes vanishingly small for $\alpha < \alpha_o$ and hence also for $\alpha < \alpha_c$. Alternatively, some additional restraint may prevent escapes altogether. If the first explanation is correct, many numerical experiments may produce a few unstable systems for $\alpha < \alpha_c < \alpha_o$. The chances of finding such systems would decrease with decreasing α. In this case, we would have to rethink our ideas of an empirical stability region and consider a more probabilistic view of hierarchical stability over the whole range of initial conditions. The second explanation seems the more likely at this stage. In the retrograde systems, the smooth trend towards an infinite stability lifetime as α decreases, implies that the region of empirical stability does exist and no unstable systems will be found within it. It is likely, though not certain, that the direct systems behave in a similar way.

It has been assumed throughout this work that $\alpha_c \leq \alpha_o$ and this assumption has been borne out by numerical experiment for the most part. An exception may be the direct systems with low ε^{23} and ε_{32} (typically both less than 10^{-5}). For these systems, the lifetime curves do not fit at all well and there is little evidence of any systematic trends (Figures 5.1). It was suggested by Walker and Roy (Paper III) that this was due to the presence of many one-spoked commensurabilities (Section 5.4). Another reason may be that there exist unstable systems with $\alpha < \alpha_c$ ($\alpha_c \to 1$ as ε^{23} , $\varepsilon_{32} \to 0$). This possibility has not been properly investigated and could be part of a larger study into the nature and value of α_o as a critical value for hierarchical stability.

8.4 Further Three-Body Numerical Experiments.

The numerical results for three-body systems presented in this thesis cover a limited range of initial conditions. In each case, the system was coplanar with initially circular osculating orbits. In addition the initial value of Θ was always zero, where Θ is the angle subtended by the Jacobian radius vectors $\underset{\sim}{\rho}_2$, $\underset{\sim}{\rho}_3$ (i.e. the system was at a conjunction of m_1 - m_2 - m_3). Most real systems do not conform to these restraints, therefore the effect of varying other parameters needs to be investigated.

The range of empirical stability parameters ε^{23} , ε_{32} was chosen to be physically relevant to the Solar System. The behaviour of the stability curves and α_o was little different for fictitious systems with ε^{23} , $\varepsilon_{32} \sim 10^{-5}$ compared to ε^{23} , $\varepsilon_{32} \sim 10^{-6}$. Thus although many real systems show ε^{23} , $\varepsilon_{32} << 10^{-6}$, it is expected that their behaviour will not differ greatly from the case when ε^{23} , $\varepsilon_{32} \simeq 10^{-6}$ (the lowest values studied). It may be of interest to perform more numerical experiments for direct systems with ε^{23} , $\varepsilon_{32} > 10^{-2}$ (the highest values studied) as they could be relevant to the study of multiple star systems.

Turning to the orbital parameters, there is the question of how the stability of a system depends on its initial configuration (Θ varying). Walker and Roy (Paper II) have studied the effects on Hill-type stability (the hierarchical preservation criterion of Chapter 3), for various initial configurations. They showed that for initially circular orbits with all other parameters constant, $\Theta = 0$ maximises stability and $\Theta = \Theta_T$ minimises it, where Θ_T varies for different systems. Another result was the following: when the system has initially eccentric orbits such that $\Theta = 0$, the most stable configuration is when the

8.4

(m_1,m_2) binary is at apocentre while the (M_2,m_3) binary is at pericentre When $\Theta = \Theta^T$, the opposite is true and the most stable configuration ariseswhen (m_1,m_2) is at pericentre and (M_2,m_3) is at apocentre. A limited number of numerical experiments based on the work of Harrington (1972) and Nacozy (1977) confirm the nature of the results. It would be of interest to perform a more detailed numerical study of the effects of varying Θ.

Most orbits in the Solar System have small eccentricities. Valsecchi, Carusi and Roy (1984) have examined the critical surfaces $\alpha_c = \alpha_c\ (\mu,\mu_3)$ for direct systems with initially eccentric orbits in a similar manner to that described in Chapter 3. They find that the surfaces are much lower than that for initially circular orbits. This implies that many of the real three-body subsystems in the Solar System have no guarantee of hierarchical preservation unless they are considered initially circular (see Section 6.4).

A similar fall in the α_c-surface was observed for retrograde systems, yet this thesis shows that the empirical region of parameters for hierarchical stability is very much larger. We are therefore encouraged to believe that a similar result may apply for systems with initially eccentric orbits.

We should also consider the case of non-planar systems and examine three-body systems with inclinations that deviate slightly from 0^o or 180^o as is the case in the Solar System. Marchal and Saari (1975) computed limits on inclinations where motion of three bodies may or may not occur. Their results could be used to limit the range of inclinations which are investigated.

All of these studies would give us a closer comparison with real

8.4

systems. However many more parameters have to be monitored and initial conditions chosen. For example in dealing with eccentric orbits, initial values for longitudes of pericentre and true anomalies must be chosen. For non-planar motion, the longitude of the ascending node must be taken into account. These studies would take considerably longer than the two cases considered in this work. However the curve-fitting procedures should be applicable.

8.5 The Four and More Body Problem

The general three-body problem still poses many unanswered questions. This is one reason why far less work has been done on the four-body problem, (the other reason almost certainly being cowardice). In considering four and more bodies, it is common to consider it as a collection of three-body subsystems which are interacting with each other.

Milani and Nobili (1985) have used this approach to study Hill-type stability in the Solar System. They consider the value of $\bar{z} = c^2 h$ which controls the topology of the forbidden regions of motion for a three-body system (Equation (3.15) in Section 3.2). If $z < z_{cr}$ then the system is hierarchically preserved, where z_{cr} is the value of z for the appropriate Euler configuration. (Note that z is negative, when the energy h is negative, whereas the quantity p/a used in Chapter 3 is positive). For an isolated three-body system, z is constant. However when various three-body subsystems interact, energy and angular momentum are exchanged; hence z is no longer constant. There is therefore no analytical Hill stability criterion for four or more bodies. It is however possible to examine the variation with time of $\Delta z(t) = z(t) - z_{cr}$. So long as $\Delta z(t) < 0$, the subsystem in question

8.5

is hierarchically preserved.

Milani and Nobili have studied $\Delta z(t)$ both by analytical means (1983) and directly from numerical integration (1985), producing useful results. For example, it is found that $\Delta z(t)$ for the Sun-Jupiter-Saturn system oscillates in anti-phase with $\Delta z(t)$ for the Sun-Uranus-Neptune system. Both quantities are negative during the time studied, implying hierarchical preservation. Moreover, there is no evidence of any secular trends in Δz which might lead to $\Delta z > 0$. Hence it implies that the Sun-Jupiter-Saturn-Uranus-Neptune system may be hierarchically preserved for the age of the Solar System.

More studies of this nature could be performed and would complement studies of temporary isolation for three-body systems discussed in Section 8.3.

Walker and Roy (Paper IV) have generalised the use of Jacobian coordinates and empirical stability parameters to n-body hierarchical systems as well as performing a more detailed study of four-body systems (Paper V). A hierarchical system may be described in terms of perturbed binary orbits. An idea of the total perturbation on a particular binary may be obtained by examining the perturbations from the other individual bodies.

For example, consider a four-body system arranged in a simple hierarchy. The equations of motion in Jacobian coordinates are given by Equation (2.16) in Section 2.4, (n = 4). The perturbations will be maximised at conjunction, when $C_{ki} = 1$, $\forall k$, $i = 2,3,4$. The perturbations on the (m_1,m_2), (M_2,m_3) and (M_3,m_4) binaries are related to Σ_2, Σ_3, Σ_4 respectively where

$$\Sigma_2 = \varepsilon_{32} + \varepsilon_{42}$$

$$\Sigma_3 = \varepsilon^{23} + \varepsilon_{43}$$

$$\Sigma_4 = \varepsilon^{24} + \varepsilon^{34}$$

these quantities are analogous to the empirical stability parameters ε^{23} , ε_{32} for a three-body system.

It is possible to perform similar numerical studies for four-body systems to those presented here for three bodies. However the number of initial conditions to be specified is larger and hence much more time and effort is necessary to acquire enough data for any kind of statistical analysis. In a paper by Roy, Walker and McDonald (1985), the results are given for several hundred experiments involving four-body systems over a limited range of parameters. The systems were assumed to be coplanar, initially circular, initially at m_1 - m_2 - m_3 - m_4 conjunction. The remaining free parameters are Σ_2, Σ_3, Σ_4, α_{23}, α_{34} ($\alpha_{24} = \alpha_{23} \cdot \alpha_{34}$, so is not independent). Fixing values of Σ_2, Σ_3, Σ_4, the initial values of α_{23} and α_{34} were varied and the resulting systems studied for hierarchical instability. In order to find how stability lifetime varies with α_{23} and α_{34}, it is necessary to fit either a two-dimensional surface or a one-dimensional curve where one of α_{23} or α_{34} is fixed and the other varied.

The studies of Milani and Nobili, and Walker and Roy, both assume that the systems in question can be broken down into three-body sub-systems. How meaningful this approach is will depend on the system chosen and whether the perturbations within the three-body subsystems are much greater than those perturbations imposed by external masses. For example, it is not especially meaningful to consider the Sun-Neptune-Pluto system as being isolated since the perturbations on Pluto's orbit

8.5

by Jupiter, Saturn and Uranus are comparable with the perturbation by Neptune (Table 6.1). In this case, we should include all five planets and the Sun in any study of Pluto's orbit.

8.6 The Search for Syzygies

In Chapter 7, an equation was derived which gives the average period of occurrence of near-syzygies for p bodies in unperturbed circular orbits about a fixed centre, (Equation 7.10). With the above assumptions, this equation can be applied to the planets of the Solar System in orbit about the (fixed) Sun or satellites in orbit about a planet. The actual numbers of near-syzygies undergone by the system can be compared with the theory. Any significant differences may indicate a commensurability in mean motions. A commensurable system may also manifest itself by showing a preference for particular types of syzygy, (eg. all planets at conjunction).

In a similar way, near-syzygies of bodies and apses can be investigated to search for resonant critical arguments. This assumes that any eccentricities are small enough to consider the mean motions as constant. It should also be possible to discover critical arguments involving the longitude of the ascending node, by considering the node as a "body". This assumes that the inclination is near enough to 0^{o} or 180^{o} for a syzygy to be meaningful. Mirror configurations can be observed by watching for near-syzygies of both body and apse for each planet of interest.

A study using this technique with a numerical integrator, was mentioned at the end of Section 7.7, and is now discussed more fully. It is applicable to the planets of the Solar System as well as existing

satellite systems. Such a study is particularly relevant at this time due to the more accurate values of the masses and positions of the outer planets and satellites being relayed by the Voyager probes. Consider a hierarchical system of p small bodies orbiting a much larger body. A numerical procedure integrates the equations of motion as accurately as possible using a large step size. A simpler method is used to interpolate between steps. For example, the orrery described in Section 7.3 could be used. Alternatively, we could use a Keplerian ellipse, updated every step. This would be more accurate but slower. Using the interpolator with small steps, near-syzygies involving any or all bodies, apses and nodes can be counted. These may then be compared with the numbers expected from Equation (7.10).

As an example consider a numerical integration for the Sun and the five outer planets. There are $\binom{5}{3} = 10$ distinct subsystems of three planets, $\binom{5}{4} = 5$ distinct subsystems of four planets and 1 system of 5 planets. There are therefore a total of 16 subsystems of three or more planets that may be studied for commensurabilities at the same time, by this technique. The same 16 subsystems still apply if we wish to investigate the occurrence of mirror configurations. However we must count syzygies of planets and apses which means the number of "bodies" is doubled compared to the previous study.

If we wish to study all possible critical arguments involving three or more apses and planets, the number of subsystems is

$$\sum_{i=3}^{10} \binom{10}{i} = 968 \text{ systems.}$$

It becomes quite a daunting task to keep a track of all possible subsystems. However many can be immediately ruled out for a number of

8.6

reasons. The apses often move so slowly between steps that they can be considered as constant and therefore the near-syzygy condition need only be considered once per step. Secondly, any system of four bodies will not be at near-syzygy if a previous subsystem of three bodies was also not at near-syzygy. Thirdly, we may rule out some systems for not being resonant on physical grounds, eg. a resonance involving the apses of Neptune and Mercury is unlikely.

It is doubtful in this example that any one would wish to study all resonances of three and more "bodies" which involve planets, apses and nodes, the total number being 32647!

When investigating a resonant system, the biggest difference between observed and expected syzygy counts will occur when the prescribed syzygy tolerance Θ is equal to twice the amplitude of oscillation of the critical argument. A priori, we do not know what this amplitude is. It is therefore advisable to search for syzygies at more than one tolerance, say $\Theta = 5^{o}$ and 20^{o}. We may then "tune into" any promising system by altering the value of Θ .

This method will not give as detailed information on a resonant system as a general perturbation theory would. It can however be used to investigate many systems simultaneously and draw attention to possible resonances that could then be analysed using more rigorous methods.

8.7 Final Remarks

This thesis has been primarily concerned with the processing of data which arises from numerical simulations of hierarchical dynamical systems. Using the methods described here, the expected behaviour of individual systems can be predicted in advance through estimates of

either the stability lifetime or the rate of occurrence of syzygies. In this way, unusual systems can be highlighted, whose behaviour is noticeably different from that predicted.

These statistical methods have been applied to studies which involve varying the fewest parameters possible, but still give useful physical results. This chapter has shown that many other studies, using different assumptions and initial conditions, are possible. At present, many of them would take years to obtain the necessary data. If however computers continue to improve in both speed and accuracy, these studies should prove less daunting in years to come. At that point, the factor which decides how long the study takes will not be the speed of the numerical integration, but rather the efficiency of the data processing.

APPENDIX A

STATISTICAL TABLES

Table 1 - Areas Under the Standard Normal Curve

Table 2 - Percentile Values for Student's t Distribution

Table 3 - Percentile Values for the Chi-Square Distribution

Table 4 - 95% Points of Skewness and Kurtosis

(Tables 1-3 are reproduced from "Mathematical Handbook of Formulas and Tables" by M. R. Spiegel)

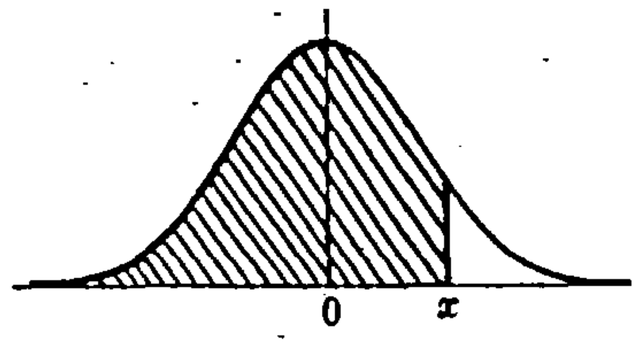
(Table 4 was computed by A.L. Brooks at the Department of Astronomy, Glasgow University).

TABLE 1

AREAS UNDER THE STANDARD NORMAL CURVE

from $-\infty$ to x

$$\operatorname{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$



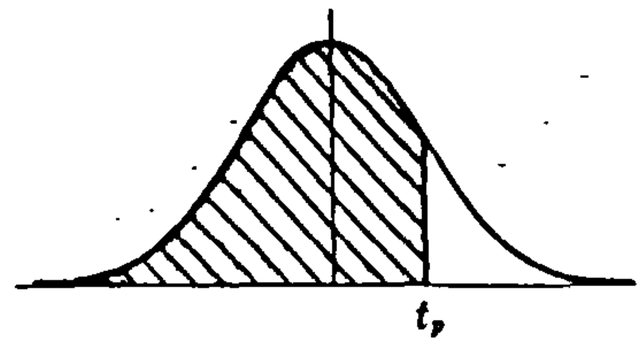
x	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5754
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7996	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 2

PERCENTILE VALUES (t_p) FOR STUDENT'S t DISTRIBUTION

with n degrees of freedom

(shaded area = p)



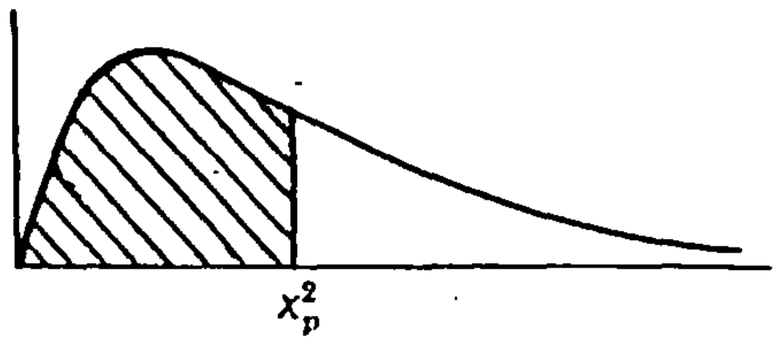
n	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.277	.137
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.03	3.36	2.57	2.02	1.48	.920	.727	.559	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.90	1.42	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.35	.870	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.692	.537	.258	.128
15	2.95	2.60	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.127
19	2.86	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.257	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.70	1.31	.855	.684	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
40	2.70	2.42	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.527	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

TABLE 3

PERCENTILE VALUES (χ_p^2) FOR THE CHI-SQUARE DISTRIBUTION

with n degrees of freedom

(shaded area $= p$)



n	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

Table 4: Simulated 95% points of skewness and kurtosis for the normal distribution.

Number of points	SKEWNESS				KURTOSIS			
	Lower point	Standard error	Upper point	Standard error	Lower point	Standard error	Upper point	Standard error
3	-0.7047	0.0001	0.7050	0.0001	1.5000	0.0000	1.5000	0.0000
4	-1.0687	0.0018	1.0685	0.0022	1.0667	0.0014	2.2996	0.0008
5	-1.2076	0.0041	1.2130	0.0035	1.2437	0.0011	3.0081	0.0030
6	-1.2390	0.0042	1.2357	0.0056	1.2906	0.0030	3.5173	0.0064
7	-1.2323	0.0056	1.2297	0.0065	1.3279	0.0024	3.8673	0.0093
8	-1.2034	0.0064	1.2080	0.0074	1.3935	0.0020	4.0985	0.0098
9	-1.1868	0.0074	1.1836	0.0060	1.4397	0.0028	4.2820	0.0141
10	-1.1652	0.0070	1.1547	0.0060	1.4795	0.0028	4.4121	0.0158
11	-1.1320	0.0063	1.1360	0.0067	1.5134	0.0026	4.5165	0.0168
12	-1.0994	0.0061	1.1145	0.0062	1.5463	0.0025	4.5457	0.0203
13	-1.0965	0.0046	1.0802	0.0059	1.5806	0.0024	4.6172	0.0196
14	-1.0585	0.0043	1.0561	0.0062	1.6044	0.0023	4.6200	0.0137
15	-1.0354	0.0064	1.0341	0.0061	1.6348	0.0023	4.6140	0.0190
16	-1.0238	0.0061	1.0176	0.0055	1.6615	0.0027	4.6494	0.0113
17	-1.0009	0.0061	0.9969	0.0068	1.6767	0.0029	4.6473	0.0166
18	-0.9684	0.0047	0.9747	0.0049	1.7003	0.0024	4.6374	0.0161
19	-0.9595	0.0048	0.9498	0.0059	1.7249	0.0028	4.6451	0.0190
20	-0.9352	0.0055	0.9291	0.0049	1.7472	0.0028	4.6280	0.0143
21	-0.9174	0.0054	0.9217	0.0054	1.7540	0.0025	4.6258	0.0150
22	-0.9022	0.0053	0.9114	0.0049	1.7764	0.0024	4.6355	0.0159
23	-0.8898	0.0053	0.8991	0.0041	1.7895	0.0030	4.6053	0.0166
24	-0.8719	0.0035	0.8691	0.0045	1.8053	0.0031	4.6122	0.0145
25	-0.8694	0.0042	0.8577	0.0049	1.8267	0.0027	4.5972	0.0145
26	-0.8480	0.0047	0.8452	0.0043	1.8378	0.0025	4.6026	0.0139
27	-0.8382	0.0044	0.8395	0.0046	1.8515	0.0029	4.5982	0.0150
28	-0.8203	0.0046	0.8284	0.0045	1.8605	0.0028	4.5817	0.0153
29	-0.8108	0.0040	0.8089	0.0053	1.8749	0.0025	4.5481	0.0170
30	-0.8015	0.0047	0.7970	0.0041	1.8896	0.0029	4.5591	0.0158
40	-0.7057	0.0040	0.7079	0.0047	1.9863	0.0026	4.4324	0.0121
50	-0.6529	0.0043	0.6438	0.0039	2.0642	0.0025	4.3569	0.0122
60	-0.5946	0.0034	0.5949	0.0029	2.1160	0.0019	4.2440	0.0111
70	-0.5528	0.0031	0.5527	0.0030	2.1684	0.0028	4.1839	0.0100
80	-0.5223	0.0033	0.5208	0.0025	2.2078	0.0025	4.1358	0.0099
90	-0.4913	0.0020	0.4896	0.0026	2.2404	0.0022	4.0789	0.0102
100	-0.4671	0.0022	0.4658	0.0022	2.2664	0.0021	4.0305	0.0105
110	-0.4462	0.0023	0.4477	0.0019	2.2948	0.0020	3.9792	0.0076
120	-0.4313	0.0022	0.4310	0.0026	2.3205	0.0021	3.9442	0.0083

APPENDIX B

LIMITING VALUES OF ε^{23} AND ε_{32}

FOR $\alpha \doteq \alpha_c$

Chapter 3 was concerned with deriving analytical criteria for the hierarchical preservation of three-body systems. The hierarchical arrangement of the bodies may be described by two binary orbits (m_1, m_2) and $(m_1 + m_2, m_3)$, where without loss of generality it is assumed that $m_2 \leq m_1$. The mass ratios μ, μ_3 are defined as

$$\mu = \frac{m_2}{m_1 + m_2}$$

$$\mu_3 = \frac{m_3}{m_1 + m_2} .$$

For given μ, μ_3, there exists a critical value for the ratio of semi-major axes of the binary orbits, denoted by α_c. For values of this ratio α which are less than α_c, the hierarchy is guaranteed to be preserved for all time.

The variables f and g are given as

$$f = \mu(1-\mu) + \frac{(1-\mu)\mu_3}{r_{13}} + \frac{\mu\mu_3}{r_{23}} \qquad (1)$$

$$g = \mu(1-\mu) + (1-\mu)\mu_3 r_{13}^2 + \mu\mu_3 r_{23}^2 \qquad (2)$$

It is found that

$$f^2 g = (1-\mu)^2\mu^2\mu_3(1+\mu_3)\alpha_c \pm 2(1-\mu)\mu\mu_3^2 (1+\mu_3)^{\frac{1}{2}} \alpha_c^{\frac{1}{2}}$$
$$+ (1-\mu)^3\mu^3(1+\mu_3) + \mu_3^3 \qquad (3)$$
$$\pm 2(1-\mu)^2\mu^2\mu_3(1+\mu_3)^{\frac{1}{2}} \alpha_c^{-\frac{1}{2}} + (1-\mu)\mu\mu_3^2 \alpha_c^{-1}$$

when $r_{13} = 1+x$, $r_{23} = x$, with x given by

$$x^5 + (3-\mu)x^4 + (3-2\mu)x^3 - (\mu+3\mu_3)x^2$$
$$-(2\mu+3\mu_3)x - (\mu+\mu_3) \doteq 0 \qquad (4)$$

corresponding to the collinear Eulerian configuration $m_1 - m_2 - m_3$. In the second and fifth terms of Equation (3), the "+" and "-" denote direct and retrograde motion respectively.

Therefore Equations (1) and (2) become

$$f = \mu(1-\mu) + \frac{(1-\mu)\mu_3}{1+x} + \frac{\mu\mu_3}{x} \qquad (5)$$

$$g = \mu(1-\mu) + (1-\mu)\mu_3(1+x)^2 + \mu\mu_3 x^2 \qquad (6)$$

By eliminating x, f, g from Equations (3), (4), (5), (6) we obtain α_c as a function of μ, μ_3.

The empirical stability parameters are

$$\begin{aligned} \varepsilon^{23} &= \mu(1-\mu)\alpha^2 \\ \varepsilon_{32} &= \mu_3 \alpha^3 \end{aligned} \qquad (7)$$

Figures 3.6 and 3.7 show the transformation of a rectangular grid of (μ, μ_3) points onto the $(\varepsilon^{23}, \varepsilon_{32})$ space when $\alpha = \alpha_c$.

Clearly there are various asymptotic limits present under this transformation which are examined in greater detail in this appendix.

$\mu = {}^1/2 \quad , \quad \mu_3 \to 0$

An example of such a system would be a binary star system with equal masses and a planet orbiting outside this binary. μ is set at its maximum value and μ_3 tends to its minimum. In the limit we are dealing with the Copenhagen problem. The limiting value of x is 0.698, obtained by solving numerically (from Equation (4)), the following polynomial.

$$x^5 + \frac{5}{2}x^4 + 2x^3 - \frac{1}{2}x^2 - x - \frac{1}{2} = 0$$

Substituting for μ in Equations (5) and (6),

$$f = \frac{1}{4} + \frac{\mu_3}{2} A$$

$$g = \frac{1}{4} + \frac{\mu_3}{2} B \qquad (8)$$

where

$$A = \frac{1}{1+x} + \frac{1}{x} \qquad (9)$$

$$B = (1+x)^2 + x^2 \qquad (10)$$

From Equations (8)

$$f^2 g = \frac{1}{64} + \frac{\mu_3}{32}(2A+B) + \frac{\mu_3^2}{16}(A^2+2AB) + \frac{\mu_3^3}{8} A^2 B \qquad (11)$$

expressed as a polynomial in μ_3. Equation (3), on substitution for μ and rearranging terms, becomes

$$f^2 g = \frac{1}{64} + \frac{\mu_3}{64}(1 + 4\alpha_c) + \frac{\mu_3^2}{16}(\alpha_c + \frac{4}{\alpha_c}) + \mu_3^3 \pm (1+\mu_3)^{\frac{1}{2}} \frac{\mu_3}{8}(\frac{1}{\alpha_c^{\frac{1}{2}}} + 4\mu_3 \alpha_c^{\frac{1}{2}}) \qquad (12)$$

Equating Equations (11) and (12), and multiplying both sides by $16/\mu_3$,

$$(A + \frac{B}{2}) + \mu_3(A^2+2AB) + \mu_3^2 \cdot 2A^2B =$$

$$\frac{1}{4} + \alpha_c + \mu_3\alpha_c + \frac{4\mu_3}{\alpha_c} + 16\mu_3^2 \pm 2(\frac{1+\mu_3}{\alpha_c^{\frac{1}{2}}})^{\frac{1}{2}}(1 + 4\mu_3\alpha_c) \qquad (13)$$

This expression is exact.

It is convenient to express α_c as a Taylor series in μ_3, since $\mu_3 \to 0$. We proceed by setting α_c equal to the leading term in the series, $c\mu_3^b$, where $b \in \mathbb{Z}$, $c \in \mathbb{R}$.

To maintain the hierarchy, α_c is constrained to lie between 0 and $1/(1-\mu) = 2$. If $b < 0$, α_c tends to $+\infty$ or $-\infty$. Thus we suppose $b = 0$, $c > 0$, i.e. $\alpha_c = c$, $\forall \mu_3$. Equating the leading terms in Equations (13), we therefore obtain

$$A + \frac{B}{2} = \frac{1}{4} + c \pm \frac{2}{c^{\frac{1}{2}}} \qquad (14)$$

By evaluating A and B at $x = 0.698$, we obtain

$$c^{3/2} - 3.46\ c^{\frac{1}{2}} \pm 2 = 0 \qquad (15)$$

(i) Direct Case

Equation (15) is a cubic polynomial in $c^{\frac{1}{2}}$.

$$(c^{\frac{1}{2}})^3 - 3.46\ c^{\frac{1}{2}} + 2 = 0 \quad .$$

There are three real solutions which yield the values

$$c = 4.412,\ 2.070,\ 0.438.$$

Of these, only $c = 0.438$ is meaningful. This implies that $\alpha_c \to 0.438$ as $\mu_3 \to 0$ for direct systems with $\mu = 0.5$. Hence $\varepsilon^{23} = \alpha_c^2/4$ tends to 0.0480. This is verified on Figure 3.6 where $\log_{10} \varepsilon^{23} \to -1.319$ as $\mu \to 0.5$ and μ_3 is small.

(ii) Retrograde Case

From Equation (15), $c = 3.46 + 2/c^{\frac{1}{2}} > 3.46$ Hence there is no physically meaningful solution for α_c with a constant leading term.

Consider now a leading term in the Taylor Series with $b = 1$, i.e. $\alpha_c = c\ \mu_3 + O(\mu_3^2)$. We substitute into Equation (13) and neglect terms of order μ_3 and higher to obtain

$$A + \frac{B}{2} = \frac{1}{4} + \frac{4}{c} - 2\frac{(1+\mu_3)^{\frac{1}{2}}}{(c\mu_3)^{\frac{1}{2}}} \qquad (16)$$

As $\mu_3 \to 0$, the right hand side of Equation (16) tends to infinity while the left hand side remains finite. Hence the leading term cannot have b = 1.

Repeating the above procedure with b = 2, i.e. $\alpha_c = c\mu_3^2 + O(\mu_3^3)$, we obtain from Equation (13)

$$A + \frac{B}{2} = \frac{1}{4} + \frac{4}{c\mu_3} - \frac{2(1+\mu_3)^{\frac{1}{2}}}{c^{\frac{1}{2}}\mu_3} + O(\mu_3)$$

Hence

$$0 = \frac{4}{c\mu_3} - \frac{2}{c^{\frac{1}{2}}\mu_3} + O(\mu_3^0)$$

which implies that c = 4. Hence $\alpha_c = 4\mu_3^2 + O(\mu_3^3)$ for retrograde systems, when $\mu = 0.5$, $\mu_3 \to 0$.

From Equations(7), with $\mu = 0.5$,

$$\log \varepsilon^{23} = \log {}^{1}\!/_{4} + 2 \log \alpha_c \qquad (17)$$

$$\log \varepsilon_{32} = \log \mu_3 + 3 \log \alpha_c \qquad (18)$$

$\alpha_c \sim 4\mu_3^2$ implies

$$\log \alpha_c = \log 4 + 2 \log \mu_3 \qquad (19)$$

Eliminating $\log \alpha_c$ and $\log \mu_3$ from Equations (17)-(19) gives

$$\log \varepsilon_{32} = \frac{7}{4} \log \varepsilon_{23} + \frac{5}{4} \log 4 \qquad (20)$$

On Figure 3.7 this equation prescribes a straight line with gradient $7/4$ and intercept at $\log \varepsilon_{32} = 0.753$.

$\mu_3 \to \infty$

The condition that μ_3 is very large pertains to the case of planet and satellite as a close binary being perturbed by the Sun. Walker

(1983) derives α_c as a series in $\mu_3^{-1/3}$ for $\mu_3 >> 1$. Although he uses a different ordering of the masses m_1, m_2, m_3 from that of Chapter 3, the techniques he employs are similar to those described in the previous case, involving careful expansion of Equations (3), (4), (5), (6) in powers of $\mu_3^{-1/3}$. He derives as the leading term in his series for α_c,

$$\alpha_c = \kappa \left(\frac{1}{81\,\mu_3}\right)^{1/3} + O\left(\frac{1}{\mu_3^{2/3}}\right) \qquad (21)$$

where κ is obtained from

$$1 \pm \frac{2}{9}\kappa^{3/2} + \frac{2}{81}\kappa^3 P_2(\cos\Theta) - \kappa = 0 \qquad (22)$$

The "+" sign in the second term of Equation (22) refers to direct motion and the "-" sign refers to retrograde motion. $P_2(x) = (3x^2-1)/2$.

Walker uses the exact expression for the energy of the three-body system. Θ is the difference in longitudes between the bodies in their orbits. However, in Chapter 3, the energy was approximated by the sum of the energies of the (m_1,m_2) and $((m_1 + m_2),m_3)$ systems. In this case the third term in Equation (22) vanishes and we are left to solve

$$1 \pm \frac{2}{9}\kappa^{3/2} - \kappa = 0 \qquad (23)$$

From Equation (21), to a first approximation,

$$\kappa = (81\,\mu_3)^{1/3}\alpha_c$$

$$= (81\,\varepsilon_{32})^{1/3} .$$

Therefore, as $\mu_3 \to \infty$, ε_{32} tends to an asymptotic limit determined by

$$1 \pm 2\,\varepsilon_{32}^{1/2} - 81^{1/3}\,\varepsilon_{32}^{1/3} = 0 \qquad (24)$$

Equation (24) is a cubic equation in $\varepsilon_{32}^{1/6}$. When solved it yields

$$\varepsilon_{32} = \begin{cases} 3.02 \times 10^{-2} & \text{for direct motion} \\ 7.10 \times 10^{-3} & \text{for retrograde motion.} \end{cases}$$

This result describes the horizontal asymptotes for $\log_{10} \varepsilon_{32} = -1.52$ and -2.15 on Figures 3.6 and 3.7 respectively. These numerical results agree with Walker's analytical results.

APPENDIX C

NUMERICAL PROCEDURE FOR ESTIMATING THE PARAMETERS OF A GAMMA DISTRIBUTION FROM A PRESCRIBED DATASET AND FOR SOLVING THE ASSOCIATED CUMULATIVE DISTRIBUTION FUNCTION

With reference to Section 5.5, we have obtained from a dataset (x,t) a subset of n points (x_i,t_i) where $|x_i - x^*| \leq \Delta$ for some values of x^* and Δ. It is proposed to fit a probability density function (p.d.f.) to the stability lifetimes $(t_i;\ i = 1,\ldots,n)$ and assume that this p.d.f. models the distribution of probable lifetimes on the interval $(x^* - \Delta,\ x^* + \Delta)$.

The p.d.f. chosen is the Gamma Distribution, namely

$$h(t) = \frac{a^b}{\Gamma(b)}\, e^{-at}\, t^{b-1} \quad , \quad (t \geq 0) \qquad (1)$$

where a and b are parameters to be determined. $\Gamma(b)$ is the Gamma Function,

$$\Gamma(b) = \int_0^\infty z^{b-1}\, e^{-z}\, dz \quad , \ (b > 0) \qquad (2)$$

For the sample (t_i), a and b must be estimated. The method that is chosen is the method of Maximum Likelihood Estimates. This well known theory can be found in detail in many books on statistics (eg. Meyer, 1978). Essentially the method is as follows.

The lifetime T is a random variable with an associated p.d.f., $h(T_i;\ a,b)$ as given in Equation (1) where T_i is a sample lifetime. Let $T_1, T_2,\ldots,T_n$ be a random sample from T and let $t_1, t_2,\ldots,t_n$ be the sample values. We define the likelihood function L as

$$L(T_1,T_2,\ldots,T_n;\ a,b) = \prod_{i=1}^{n} h(T_i;\ a,b) \qquad (3)$$

(This is merely the joint p.d.f. for a set of independent random variables T_i with the same p.d.f.). It is assumed that we have measured the sample values $t_1,\ldots,t_n$, but we must still determine a and b. The

question that is asked is, "What values of a,b are most likely to have produced a sample of values $t_1,\ldots,t_n$?" The definition of the maximum likelihood estimates of a and b are those values that maximise L for the given sample. log L will achieve its maximum value for the same values of a,b as will L and in general this is an easier function to use. We therefore have to solve the following equations simultaneously.

$$\frac{\partial}{\partial a} (\log L (t_1,t_2,\ldots,t_n;a,b)) = 0 \qquad (4)$$

$$\frac{\partial}{\partial b} (\log L (t_1,t_2,\ldots,t_n;a,b)) = 0 \qquad (5)$$

The application of maximum likelihood estimates to the Gamma Distribution is described by Chapman (1953). Substituting Equation (1) into Equation (3) gives

$$L(t_1,\ldots,t_n;a,b) = \prod_{i=1}^{n} \frac{a^b}{\Gamma(b)} e^{-at_i} t_i^{b-1} \qquad (6)$$

from which it can be shown that

$$\frac{1}{n} \log L = b \log a - \log \Gamma(b) - a\bar{t} + (b-1) \bar{t}_L \qquad (7)$$

where

$$n\bar{t} = \sum_{i=1}^{n} t_i \qquad (8)$$

$$n\bar{t}_L = \sum_{i=1}^{n} \log t_i \qquad (9)$$

On substitution of Equation (7), Equations (4) and (5) become

$$\frac{b}{a} - \bar{t} = 0 \qquad (10)$$

$$\log a - \frac{\Gamma'(b)}{\Gamma(b)} + \bar{t}_L = 0 \qquad (11)$$

Thus from Equation (10)

$$a = b/\bar{t} \tag{12}$$

which when substituted into Equation (11) gives

$$\phi(b) = \log b - \frac{\Gamma'(b)}{\Gamma(b)} - c = 0 \tag{13}$$

where

$$c = \log \bar{t} - \bar{t}_L \tag{14}$$

is known. There is no analytical solution to Equation (13) for b. Tables are available which give numerically derived values for b, given c, (Chapman, 1953). It is however relatively simple to obtain a value of b to arbitrary accuracy for any given c, by the following method.

We use the analytical result (Spiegel, 1968)

$$\frac{\Gamma'(b)}{\Gamma(b)} = -\gamma + \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{b+k-1} \right)$$

$$= -\gamma + (b-1) \sum_{k=1}^{\infty} \frac{1}{k(b+k-1)} \tag{15}$$

in Equation (13) to give

$$\phi(b) = \log b + (\gamma - c) - (b-1) \sum_{k=1}^{\infty} \frac{1}{k(b+k-1)} \tag{16}$$

where γ is Euler's constant ($\gamma = 0.5772156....$) .

After a little algebra, it can be seen that the derivative of $\phi(b)$ is

$$\phi'(b) = \frac{1}{b} - \sum_{k=1}^{\infty} \frac{1}{(b+k-1)^2} \tag{17}$$

We are then able to use a Newton-Raphson iterative scheme to solve $\phi(b) = 0$, where each successive approximation to b is

$$b_1 = b - \frac{\phi(b)}{\phi'(b)} \tag{18}$$

The procedure to find the best estimates of a,b is as follows:-

(i) From the dataset (t_i), compute c from Equations (8), (9) and (14).

(ii) Solve Equation (13) for b by the Newton-Raphson technique using Equations (16), (17) and (18). Obviously the infinite series must be truncated at some value of k, when the summation is sufficiently accurate.

(iii) Compute a from Equation (12).

In this way we have determined the most likely probability distribution given the dataset we possess. In order to make use of it, (and test its validity) this p.d.f. should allow us to predict with quantitative uncertainty what values further measurements are likely to take. Put another way, if we run 100 test systems and note when they become unstable how long must we wait before 10, 25, 50, 75 or 90 become unstable?

To answer this question, we must study the cumulative distribution function (c.d.f.)

$$H(T) = \int_0^T h(t)\,dt \qquad (T > 0) \qquad (19)$$

H(T) gives the probability of a random sample having value $t \leq T$.

In our usage, it gives the probability that a system will become unstable within time T. Note that $H(0) = 0$, $H(T) \to 1$ as $T \to \infty$.

From Equations (1) and (19),

$$H(T;a,b) = \int_0^T \frac{a^b}{\Gamma(b)}\, e^{-at}\, t^{b-1} \qquad (T \geq 0) \qquad (20)$$

Let $u = at$ and $U = aT$. Then Equation (20) becomes

$$H(b,U) = \frac{1}{\Gamma(b)} \int_0^U e^{-u}\, u^{b-1}\, du \qquad (21)$$

For a chosen probability $H(b,U)$, Equation (21) must be solved for U.

Define $\Gamma(b,U)$ by

$$\Gamma(b,U) = \Gamma(b)\,(1 - H(b,U)) \qquad (22)$$

We make use of the analytical result, (Abramowitz and Stegun, 1972),

$$\Gamma(b,U) = e^{-U}\, U^b\, C(b,U) \qquad (23)$$

where $C(b,U)$ is a continued fraction.

$$C(b,U) = \frac{1}{U+}\ \frac{(1-b)}{1+}\ \frac{1}{U+}\ \frac{(2-b)}{1+}\ \frac{2}{U+}\ \ldots. \qquad (24)$$

$\Gamma(b)$ may be approximated by a power series

$$\Gamma(b+1) = \sum_{i=0}^{\infty} \sigma_i\, b^i \qquad (0 \leq b \leq 1) \qquad (25)$$

making use of the result

$$\Gamma(b+1) = b\Gamma(b) \qquad (26)$$

to reduce b to the required range for Equation (25), if necessary.

The first few coefficients σ_i are given by Abramowitz and Stegun (1972) as

$$\begin{aligned}
\sigma_0 &= 1 \\
\sigma_1 &= -0.5771917 \\
\sigma_2 &= 0.9882059 \\
\sigma_3 &= -0.8970569 \\
\sigma_4 &= 0.9182069 \\
\sigma_5 &= -0.7567041 \\
\sigma_6 &= 0.4821994 \\
\sigma_7 &= -0.1935278 \\
\sigma_8 &= 0.0358683
\end{aligned} \qquad (27)$$

From Equation (22),

$$\Gamma(b,U) - K = 0 \qquad (28)$$

where

$$K = \Gamma(b)\ (1 - H(b,U)) \qquad (29)$$

is a constant (remember that H(b,U) is chosen).

Once again we use the Newton-Raphson method, this time to solve Equation (28) for U. Note that

$$\frac{d}{dU}\Gamma(b,U) = \frac{d}{dU}\left[\Gamma(b) - \Gamma(b)\,H(b,U)\right]$$

$$= -\frac{d}{dU}\int_0^U e^{-u}\,u^{b-1}\,du$$

$$= -e^{-U}\,U^{b-1}$$

Hence the Newton-Raphson iteration uses

$$U_1 = U + \frac{(\Gamma(b,U) - K)}{e^{-U}\,U^{b-1}}$$

i.e. $$U_1 = U\,(1 + C(b,U) - K\,e^{U}\,U^{-b}) \qquad (30)$$

from Equation (23).

The procedure to find U is as follows:-

(i) Set H(b,U) equal to a chosen probability level.

(ii) Calculate $\Gamma(b)$ from Equations (25), (26) and (27).

(iii) Solve Equation (28) for U by the Newton-Raphson technique using Equations (24), (29) and (30). The continued fraction C(b,U) must be truncated when sufficiently accurate.

In this way, for any probability, a lifetime T = U/a may be calculated such that the corresponding proportion of systems will have stability lifetimes less than T.

REFERENCES

Abramowitz M., & Stegun I.A.: 1972, Handbook of Mathematical Functions, Dover.

Aksnes, K., 1985: The Tiny Satellites of Jupiter and Saturn and their Interaction with the Rings. in Stability of the Solar System and its Minor Natural and Artificial Bodies, (Reidel).

American Ephemeris and Nautical Almanac, 1960. U.S.Government.

Anderson T.W., & Sclove S.L., 1978: An Introduction to the Statistical Analysis of Data. Houghton Mifflin.

Arnol'd V.I., 1963: Usp. Mat. Nauk (Sov. Math-Usp) 18, No.5, 13: No.6,91.

Astronomical Almanac, 1986. H.M.S.O.

Birkhoff G.D., 1927: Dynamical Systems, (American Mathematical Society Providence, R.I.).

Boyle C.B., 1984: Ph.D. Thesis.

Brouwer, D. 1937: On the Accumulation of Errors in Numerical Integration, Astronomical Journal 46, 149.

Brouwer D. 1963: Astronomical Journal 68, 152.

Brown E.W., 1896: An Introductory Treatise on the Lunar Theory. Cambridge University Press.

Cambridge Atlas of Astronomy, 1985: eds. J. Audouze & G. Israel. Newnes Books.

Chapman D.G., 1956: Estimating the Parameters of a Truncated Gamma Distribution. Annals of Mathematical Statistics 27, 498.

Cohen C.J., & Hubbard E.C., 1965: Astron.,J. 70, 10.

Cohen C.J., Hubbard E.C., Oesterwinter C., 1967: Astron. J., 72, 973.

Cohen C.J., Hubbard E.C., Oesterwinter C., 1972: Astron. Pap. Am. Ephemeris 13.

Cohen C.J., Hubbard E.C., Oesterwinter C., 1973: Cel. Mech. 7, 438.

Cole G.H.A., 1985: Dynamical Form of the Solar System. Observatory, Vol.105, No.1066, 96.

Connaissance Des Temps. 1970: Le Bureau Des Longitudes, Gautier-Villars.

Contopoulos G. 1983a: Infinite Bifurcations, Gaps and Bubbles in Hamiltonian Systems. Physica 8D, 142.

Contopoulos G., 1983b: Theoretical Periodic Orbits in 3-Dimensional Hamiltonians, Physica 11D, 179.

Contopoulos G., 1983c: Termination of Sequences of Bifurcations in 3-Dimensional Hamiltonian Systems. Lettre al Nuovo Cimento, 38, 7.

Contopoulos G., & Pinotsis A., 1984: Infinite Bifurcations in the Restricted 3-Body Problem, Astron. Astrophys., 133, 49.

Danby J.M.A., 1962: Fundamentals of Celestial Mechanics, Macmillan.

Evans D.S., 1968: Quarterly J. Roy. Astr. Soc., 9, 388.

Explanatory Supplement to the Astronomical Ephemeris 1961: H.M.S.O.

Fox K.: 1982: Thesis, Appendix A. Numerical Integration of the Equations of Motion of Meteoroids.

Goldreich P., 1965: An Explanation of the Frequent Occurrence of Commensurable Mean Motions in the Solar System. M.N.R.A.S. 130, 159.

Goudas C.L., & Petsagourakis E.G., 1985: Motions in the Magnetic Field of Two Revolving Dipoles. Stability of the Solar System and its Minor Natural and Artificial Bodies (Reidel).

Hagihara Y., 1952: Proc. Japan Acad. 28, 182.

Hagihara Y., 1957: Stability in Celestial Mechanics (Kasai, Tokyo).

Harrington R.S., 1972: Stability Criteria for Triple Systems. Cel. Mech. 6, 322.

Harrington R.S., 1977: Planetary Orbits in Binary Stars. Astron. Jour. 82 (9), 753.

Hénon M. 1966: Bull. Astron. 1, 57.

Text cut off in original

Henon M., 1970: Astron. Astrophys. 9, 24.

Hill G., 1878: Am. J. Math. 1, 129, 245.

Hodge S.E. & Seed M.L., 1972: Statistics and Probability, Blackie Chambers.

Jacobi C.G.J., 1836: Compte Rendus de l'Acad. des Sciences III, 59, Paris.

Jefferys W.H., & Szebehely V.G., 1978: Comments on Astrophysics 8(1),9.

Isaacson E, & Keller H.B. 1966: Analysis of Numerical Methods, John Wiley & Son

Kolmogorov A.N., 1954: Doklady Akad. Nauk. SSSR (Sov. Phys - Doklady),98 No.4

Kinoshita H. & Nakai H. 1984: Celest. Mech., 34, 203.

Lazzaro D., Ferraz-Mello S., & Veillet C., 1984: The Laplacian Resonance Amongst Uranian Inner Satellites. Astron. Astrophys. 140, 33.

Mandelbrot B.B., 1982: The Fractal Geometry of Nature. (Oxford, Freeman).

Marchal C., 1985: Three-Body Problem: Some Applications of the Tests of Escape. Stability of the Solar System and its Minor Natural and Artificial Bodies.

Marchal C., & Saari D.G., 1975: Hill Regions for the General Three-Body Problem, Celest. Mech. 12, 115.

Markellos V.V., 1973: Ph.D. Thesis.

Markellos V.V., 1974a: Numerical Investigation of the Planar Restricted Three-Body Problem. I. Periodic Orbits of the Second Generation in the Sun/Jupiter System. Cel. Mech. 9, 365.

Markellos V.V., 1974b: Numerical Investigation of the Planar Restricted Three-Body Problem. II. Regions of Stability for Retrograde Satellites of Jupiter as determined by Periodic Orbits of the Second Generation. Cel. Mech. 10, 87.

Message P.J., 1966: I.A.U. Symposium No.25, p.197. (New York, Academic).

Message P.J., 1978: Proceedings of I.A.U. Symposium No.41, Reidel, Dordrecht.

Meyer P.L., 1978: Introductory Probability and Statistical Applications. Addison-Wesley.

Mignard F., 1982: Radiation Pressure and Dust Particle Dynamics. Icarus 49, 347.

Mignard F., & Henon M., 1984: About an Unsuspected Integrable Problem. Celest. Mech., 33, 239.

Milani A., & Nobili A.M., 1983: On the Stability of Hierarchical Four-Body Systems. Celest. Mech., 31, 241.

Milani A., & Nobili A.M., 1984: Resonance Locking between Jupiter and Uranus. Nature 310, 135.

Milani A., & Nobili A.M., 1985: Methods of Stability Analysis in the Solar System. Stability of the Solar System and its Minor Natural and Artificial Bodies.

Moser J., 1973: Stable and Random Motions in Dynamical Systems. Princeton University Press, Princeton, N.J.

Nacozy P.E., 1977: A Discussion of Long-term Numerical Solutions of the Jupiter-Saturn-Sun System. Celest.Mech., 16, 77.

Newcomb S., 1876: Smithsonian Contribution to Knowledge 21.

Ovenden M.W., & Roy A.E.: 1960: Mon. Not. R. Astron. Soc., 123. 1.

Plummer H.C., 1918: An Introductory Treatise on Dynamical Astronomy, Cambridge University Press.

Poincaré H., 1895: Les Méthodes Nouvelles de la Méchanique Célest. Gauthier-Villars, Paris.

Roy A.E., 1979: Orbital Motion (2nd Edition). Adam Hilger.

Roy A.E., 1972: Commensurable Mean Motions as a Tool in Solar System Dynamical Studies. Proceedings of the First European Astronomical Meeting, Athens, September 4-9, 1972. Vol.2.

Roy A.E., 1983: Project Longstop. Private Memorandum.

Roy A.E., & Ovenden M.W.: 1954: On the Occurrence of Commensurable Mean Motions in the Solar System. M.N.R.A.S. 114, 232.

Roy A.E., & Ovenden M.W., 1955: On the Occurrence of Commensurable Mean Motions in the Solar System. II. The Mirror Theorem. M.N.R.A.S. 115, 298

Roy A.E., Walker I.W., & McDonald A.J.C., 1985: Studies in the Stability of Hierarchical Dynamical Systems . Stability of the Solar System and its Minor Natural and Artificial Bodies.

Sarris E., 1982: Integrals of Motion in the Elliptic Three-Dimensional Restricted Three-Body Problem. Celest. Mech. 26, 353.

Schuerman D.W., 1980: The Restricted Three-Body Problem with Radiation Pressure. Ap.J. Vol.238, 337.

Schwarz H.E. & Walker I.W.: 1982: Studies in the Application of Recurrence Relations to Specific Perturbation Methods. VI. Comparison with Classical Single-Step and Multi-Step Methods of Numerical Integration.

Simmons J.F.L., McDonald A.J.C. & Brown J.C., 1985: The Restricted 3-Body Problem with Radiation Pressure. Celestial Mechanics 35, 145.

Spiegel M.R., 1968: Schaum's Outline Series in Mathematics – Mathematical Handbook of Formulas and Tables. McGraw-Hill.

Spirig F., & Waldvogel J., 1985: The Three-Body Problem with Two Small Masses: A Singular-Perturbation Approach to the Problem of Saturn's Coorbiting Satellites. Stability of the Solar System and its Minor Natural and Artificial Bodies. (Reidel).

Szebehely V., 1977: Analytical Determination of the Measure of Stability of Triple Stellar Systems. Celest. Mech. 15, 107.

Szebehely V.& McKenzie R., 1977a: Stability of Planetary Systems with Bifurcation Theory. Astron. Jour., 82(1), 79.

Szebehely V. & McKenzie R.: 1977b: Stability of the Sun-Earth-Moon System. Astron. Jour. 82(4), 303.

Szebehely V. & Zare K., 1977: Stability of Classical Triplets and of their Hierarchy. Astron. Astrophys. 58, 145.

Valsecchi G.B., Carusi A., & Roy A.E., 1984 : The Effect of Orbital Eccentricities on the Shape of the Hill-type Analytical Stability Surfaces in the General Three-Body Problem. Celest. Mech. 32, 217.

Walker I.W., 1980: Ph.D. Thesis. Empirical Stability Criteria for Hierarchical Many-Body Dynamical Systems in Celestial Mechanics.

Walker I.W., 1983: On the Stability of Close Binaries in Hierarchical Three-Body Systems. Celest. Mech. 29, 215.

Walker I.W., Emslie A.G., & Roy A.E., 1980: Stability Criteria in Many-Body Systems. I. An Empirical Stability Criterion for Co-rotational Three-Body Systems. Celest. Mech. 22, 371.

Walker I.W. & Roy A.E., 1981: Stability Criteria in Many-Body Systems. II. On a Sufficient Condition for the Stability of Coplanar Hierarchical Three-Body Systems. Celest. Mech. 24, 195.

Walker I.W. & Roy A.E., 1983: Stability Criteria in Many-Body Systems. III. Empirical Stability Regions for Corotational, Coplanar, Hierarchical Three-Body Systems. Celest. Mech. 29, 117.

Walker I.W., 1983: Stability Criteria in Many-Body Systems. IV. Empirical Stability Parameters for General Hierarchical Dynamical Systems. Celest. Mech. 29, 149.

Walker I.W. & Roy A.E., 1983: Stability Criteria in Many-Body Systems. V. On the Totality of Possible Hierarchical General Four-Body Systems. Celest. Mech. 29, 267.

Wiesel W., 1980: Orbital and Solar Resonance in the Jovian Moon System. Celest. Mech. 21, 265.

Williams J.G., & Benson G.S., 1971: Astron. J. 76, 167.

Zare K., 1976: The Effects of Integrals on the Totality of Solutions of Dynamical Systems. Celest. Mech. 14, 73.

Zare K., 1977: Bifurcation Points in the Planar Problem of Three Bodies. Celest. Mech. 16, 35.