7
Secular Perturbations

Past and to come seem best, things present worst.

William Shakespeare, *Henry IV, (2), I, iii*

7.1 Introduction

In the last chapter we saw how the disturbing function can be expanded in an infinite series where the individual terms can be classified as secular, resonant, or short period, according to the given physical problem. We have already stated in Sect. 3 that the N-body problem (for $N \geq 3$) is nonintegrable. However, in this chapter we will show how, with suitable approximations, it is possible to find an analytical solution to a particular form of the N-body problem that can be applied to the motion of solar system bodies. We can do this by considering the effects of the purely secular terms in the disturbing function for a system of N masses orbiting a central body. The resulting theory can be applied to satellites orbiting a planet, or planets orbiting the Sun, and then used to study the motion of small objects orbiting in either of these systems. This is the subject of *secular perturbation theory.*

7.2 Secular Perturbations for Two Planets

Consider the motion of two planets of mass m_1 and m_2 moving under their mutual gravitational effects and the attraction of a point-mass central body of mass m_c where $m_1 \ll m_c$ and $m_2 \ll m_c$. Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be the disturbing functions describing the perturbations on the orbit of the masses m_1 and m_2 respectively, where $\mathcal{R}_1$ and $\mathcal{R}_2$ are functions of the standard *osculating* orbital elements of both bodies. Osculating elements, from the Latin verb *osculare* meaning "to kiss", are instantaneous elements derived from the values of the position and velocity of an object assuming an unperturbed keplerian orbit. The perturbations on the orbital elements are given by Lagrange's equations, Eqs. (6.145)–(6.150).

In the absence of any mean motion commensurabilities between two masses, the secular perturbations arising from the gravitational perturbations between m_1, m_2, and m_c are obtained by isolating the terms in the disturbing function that are independent of the mean longitudes. We can also exclude any terms that depend only on the semi-major axis since, from Eq. (6.145), these will not make any contribution to secular evolution. To second order in the eccentricities and inclinations (and first order in the masses), the only terms in the expansion of the disturbing function that do not contain the mean longitudes are, from Appendix B, the terms 4D0.1, 4D0.2, and 4D0.3 with $j = 0$. Hence the general, averaged, secular, direct part of the disturbing function is

$$\begin{aligned}\mathcal{R}_{\mathrm{D}}^{(\mathrm{sec})} = & \frac{1}{8}\left[2\alpha_{12}D + \alpha_{12}^2 D^2\right] b_{\frac{1}{2}}^{(0)}\,(e_1^2 + e_2^2) - \frac{1}{2}\alpha_{12} b_{3/2}^{(1)}(s_1^2 + s_2^2) \\ & + \frac{1}{4}\left[2 - 2\alpha_{12}D - \alpha_{12}^2 D^2\right] b_{\frac{1}{2}}^{(1)}\, e_1 e_2 \cos(\varpi_1 - \varpi_2) \\ & + \alpha_{12} b_{3/2}^{(1)} s_1 s_2 \cos(\Omega_1 - \Omega_2), \end{aligned} \tag{7.1}$$

where the subscripts 1 and 2 refer to the inner and outer body respectively and $\alpha_{12} = a_1/a_2$ where $a_1 < a_2$. There is no indirect part. In fact, as can be seen from Appendix B, all the indirect terms involve at least one mean longitude and hence will never contribute purely secular terms (see Brouwer & Clemence 1961).

When calculating $\mathcal{R}_1$ and $\mathcal{R}_2$ from $\mathcal{R}_{\mathrm{D}}^{(\mathrm{sec})}$ we have to take account of the fact that $\mathcal{R}_1$ arises from an external perturbation by m_2 whereas $\mathcal{R}_2$ comes from an internal perturbation by m_1. Hence, from Eqs. (6.134) and (6.135), $\mathcal{R}_1$ and $\mathcal{R}_2$ can be written as

$$\mathcal{R}_1 = \frac{\mathcal{G}m_2}{a_2}\mathcal{R}_{\mathrm{D}}^{(\mathrm{sec})} = \frac{\mathcal{G}m_2}{a_1}\alpha_{12}\mathcal{R}_{\mathrm{D}}^{(\mathrm{sec})} \tag{7.2}$$

and

$$\mathcal{R}_2 = \frac{\mathcal{G}m_1}{a_1}\alpha_{12}\mathcal{R}_{\mathrm{D}}^{(\mathrm{sec})} = \frac{\mathcal{G}m_1}{a_2}\mathcal{R}_{\mathrm{D}}^{(\mathrm{sec})}. \tag{7.3}$$

Using the following relationships between the Laplace coefficients and their derivatives,

$$2\alpha\frac{\mathrm{d}b_{1/2}^{(0)}}{\mathrm{d}\alpha} + \alpha^2\frac{\mathrm{d}^2 b_{1/2}^{(0)}}{\mathrm{d}\alpha^2} = \alpha b_{3/2}^{(1)}, \tag{7.4}$$

$$2b_{1/2}^{(1)} - 2\alpha\frac{\mathrm{d}b_{1/2}^{(1)}}{\mathrm{d}\alpha} - \alpha^2\frac{\mathrm{d}^2 b_{1/2}^{(1)}}{\mathrm{d}\alpha^2} = -\alpha b_{3/2}^{(2)}, \tag{7.5}$$

and the approximations $\mathcal{G}m_c \approx n_1^2 a_1^3 \approx n_2^2 a_2^3$, we can write

$$\mathcal{R}_1 = n_1^2 a_1^2 \frac{m_2}{m_c + m_1}\left[\frac{1}{8}\alpha_{12}^2 b_{3/2}^{(1)} e_1^2 - \frac{1}{8}\alpha_{12}^2 b_{3/2}^{(1)} I_1^2 - \frac{1}{4}\alpha_{12}^2 b_{3/2}^{(2)} e_1 e_2 \cos(\varpi_1 - \varpi_2) + \frac{1}{4}\alpha_{12}^2 b_{3/2}^{(1)} I_1 I_2 \cos(\Omega_1 - \Omega_2)\right] \tag{7.6}$$

and

$$\mathcal{R}_2 = n_2^2 a_2^2 \frac{m_1}{m_c + m_2}\left[\frac{1}{8}\alpha_{12} b_{3/2}^{(1)} e_2^2 - \frac{1}{8}\alpha_{12} b_{3/2}^{(1)} I_2^2 - \frac{1}{4}\alpha_{12} b_{3/2}^{(2)} e_1 e_2 \cos(\varpi_1 - \varpi_2) + \frac{1}{4}\alpha_{12} b_{3/2}^{(1)} I_1 I_2 \cos(\Omega_1 - \Omega_2)\right], \tag{7.7}$$

where we have assumed that I_1 and I_2 are small enough so that the approximations $s_1 = \sin\frac{1}{2}I_1 \approx \frac{1}{2}I_1$ and $s_2 = \sin\frac{1}{2}I_2 \approx \frac{1}{2}I_2$ are valid.

The equations for $\mathcal{R}_1$ and $\mathcal{R}_2$ given in Eqs. (7.6) and (7.7) can be combined to give

$$\mathcal{R}_j = n_j a_j^2\left[\frac{1}{2}A_{jj} e_j^2 + A_{jk} e_1 e_2 \cos(\varpi_1 - \varpi_2) \frac{1}{2}B_{jj} I_j^2 + B_{jk} I_1 I_2 \cos(\Omega_1 - \Omega_2)\right], \tag{7.8}$$

where $j = 1, 2$; $k = 2, 1$ ($j \neq k$); and

$$A_{jj} = +n_j \frac{1}{4}\frac{m_k}{m_c + m_j}\alpha_{12}\bar{\alpha}_{12} b_{3/2}^{(1)}(\alpha_{12}), \tag{7.9}$$

$$A_{jk} = -n_j \frac{1}{4}\frac{m_k}{m_c + m_j}\alpha_{12}\bar{\alpha}_{12} b_{3/2}^{(2)}(\alpha_{12}), \tag{7.10}$$

$$B_{jj} = -n_j \frac{1}{4}\frac{m_k}{m_c + m_j}\alpha_{12}\bar{\alpha}_{12} b_{3/2}^{(1)}(\alpha_{12}), \tag{7.11}$$

$$B_{jk} = +n_j \frac{1}{4}\frac{m_k}{m_c + m_j}\alpha_{12}\bar{\alpha}_{12} b_{3/2}^{(1)}(\alpha_{12}), \tag{7.12}$$

where $\bar{\alpha}_{12} = \alpha_{12}$ if $j = 1$ (an external perturber) and $\bar{\alpha}_{12} = 1$ if $j = 2$ (an internal perturber). From the definition of the Laplace coefficients given in Sect. 6.4 we have

$$b_{3/2}^{(1)}(\alpha) = \frac{1}{\pi}\int_0^{2\pi} \frac{\cos\psi \, \mathrm{d}\psi}{(1 - 2\alpha\cos\psi + \alpha^2)^{\frac{3}{2}}}, \tag{7.13}$$

$$b_{3/2}^{(2)}(\alpha) = \frac{1}{\pi}\int_0^{2\pi} \frac{\cos 2\psi \, \mathrm{d}\psi}{(1 - 2\alpha\cos\psi + \alpha^2)^{\frac{3}{2}}}. \tag{7.14}$$

Note that in this case $B_{11} = -B_{12}$ and $B_{21} = -B_{22}$. However, the situation is different when we have to take account of terms due to the oblateness of the central body (see Sect. 7.7). All these quantities are frequencies that can be thought of as the constant elements of two matrices **A** and **B** given by

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}. \tag{7.15}$$

Note that the elements of these matrices are only functions of the masses and the (fixed) semi-major axes of the two bodies and that the rows (or columns) of the matrix **B** are not linearly independent.

Taking the lowest order terms in e and I in Eqs. (6.146), (6.148), (6.149), and (6.150) we can easily derive an approximate form of Lagrange's equations for the time variation of the original orbital elements:

$$\dot{e}_j = -\frac{1}{n_j a_j^2 e_j}\frac{\partial \mathcal{R}_j}{\partial \varpi_j}, \qquad \dot{\varpi}_j = +\frac{1}{n_j a_j^2 e_j}\frac{\partial \mathcal{R}_j}{\partial e_j}, \tag{7.16}$$

$$\dot{I}_j = -\frac{1}{n_j a_j^2 I_j}\frac{\partial \mathcal{R}_j}{\partial \Omega_j}, \qquad \dot{\Omega}_j = +\frac{1}{n_j a_j^2 I_j}\frac{\partial \mathcal{R}_j}{\partial I_j}. \tag{7.17}$$

Given the form of the equations above, it is convenient to define the vertical and horizontal components of eccentricity and inclination "vectors" by:

$$h_j = e_j \sin \varpi_j, \qquad k_j = e_j \cos \varpi_j \tag{7.18}$$

and

$$p_j = I_j \sin \Omega_j, \qquad q_j = I_j \cos \Omega_j. \tag{7.19}$$

These variables have the advantage that they avoid the singularities inherent in Eqs. (7.16) and (7.17) for low e and I. The general secular part of the disturbing function can now be written

$$\mathcal{R}_j = n_j a_j^2 \Big[\frac{1}{2} A_{jj}(h_j^2 + k_j^2) + A_{jk}(h_j h_k + k_j k_k) + \frac{1}{2} B_{jj}(p_j^2 + q_j^2) + B_{jk}(p_j p_k + q_j q_k) \Big]. \tag{7.20}$$

Note that when k is used as a subscript it is always equal to either 1 or 2, denoting the interior or exterior body; this should not be confused with the use of k as the horizontal component of the eccentricity vector.

Since each of the h_j, k_j, p_j, and q_j is a function of two variables we can write

$$\frac{\mathrm{d}h_j}{\mathrm{d}t} = \frac{\partial h_j}{\partial e_j}\frac{\mathrm{d}e_j}{\mathrm{d}t} + \frac{\partial h_j}{\partial \varpi_j}\frac{\mathrm{d}\varpi_j}{\mathrm{d}t}, \qquad \frac{\mathrm{d}k_j}{\mathrm{d}t} = \frac{\partial k_j}{\partial e_j}\frac{\mathrm{d}e_j}{\mathrm{d}t} + \frac{\partial k_j}{\partial \varpi_j}\frac{\mathrm{d}\varpi_j}{\mathrm{d}t}, \tag{7.21}$$

$$\frac{\mathrm{d}p_j}{\mathrm{d}t} = \frac{\partial p_j}{\partial I_j}\frac{\mathrm{d}I_j}{\mathrm{d}t} + \frac{\partial p_j}{\partial \Omega_j}\frac{\mathrm{d}\Omega_j}{\mathrm{d}t}, \qquad \frac{\mathrm{d}q_j}{\mathrm{d}t} = \frac{\partial q_j}{\partial I_j}\frac{\mathrm{d}I_j}{\mathrm{d}t} + \frac{\partial q_j}{\partial \Omega_j}\frac{\mathrm{d}\Omega_j}{\mathrm{d}t}, \tag{7.22}$$

where, from the definitions given above, the partial derivatives are given by

$$\frac{\partial h_j}{\partial e_j} = \frac{h_j}{e_j}, \qquad \frac{\partial k_j}{\partial e_j} = \frac{k_j}{e_j}, \qquad \frac{\partial h_j}{\partial \varpi_j} = +k_j, \qquad \frac{\partial k_j}{\partial \varpi_j} = -h_j \tag{7.23}$$

and

$$\frac{\partial p_j}{\partial I_j} = \frac{p_j}{I_j}, \qquad \frac{\partial q_j}{\partial I_j} = \frac{q_j}{I_j}, \qquad \frac{\partial p_j}{\partial \Omega_j} = +q_j, \qquad \frac{\partial q_j}{\partial \Omega_j} = -p_j. \tag{7.24}$$

After some calculation it can be shown that the perturbation equations can be written as

$$\dot{h}_j = +\frac{1}{n_j a_j^2}\frac{\partial \mathcal{R}_j}{\partial k_j}, \qquad \dot{k}_j = -\frac{1}{n_j a_j^2}\frac{\partial \mathcal{R}_j}{\partial h_j}, \tag{7.25}$$

$$\dot{p}_j = +\frac{1}{n_j a_j^2}\frac{\partial \mathcal{R}_j}{\partial q_j}, \qquad \dot{q}_j = -\frac{1}{n_j a_j^2}\frac{\partial \mathcal{R}_j}{\partial p_j}, \tag{7.26}$$

where $\mathcal{R}_j$ is as given in Eq. (7.20).

The full equations for the variation of h_j, k_j, p_j, and q_j ($j = 1, 2$) then become

$$\begin{aligned}
\dot{h}_1 &= +A_{11}k_1 + A_{12}k_2, & \dot{k}_1 &= -A_{11}h_1 - A_{12}h_2,\\
\dot{h}_2 &= +A_{21}k_1 + A_{22}k_2, & \dot{k}_2 &= -A_{21}h_1 - A_{22}h_2,\\
\dot{p}_1 &= +B_{11}q_1 + B_{12}q_2, & \dot{q}_1 &= -B_{11}p_1 - B_{12}p_2,\\
\dot{p}_2 &= +B_{21}q_1 + B_{22}q_2, & \dot{q}_2 &= -B_{21}p_1 - B_{22}p_2.
\end{aligned} \tag{7.27}$$

Thus, to lowest order, the equations for the time variation of $\{h_j, k_j\}$ are decoupled from those of $\{p_j, q_j\}$. Furthermore, these are linear differential equations with constant coefficients, and hence the problem of secular perturbations reduces to two sets of eigenvalue problems. The solutions are given by

$$h_j = \sum_{i=1}^{2} e_{ji} \sin(g_i t + \beta_i), \qquad k_j = \sum_{i=1}^{2} e_{ji} \cos(g_i t + \beta_i), \tag{7.28}$$

$$p_j = \sum_{i=1}^{2} I_{ji} \sin(f_i t + \gamma_i), \qquad q_j = \sum_{i=1}^{2} I_{ji} \cos(f_i t + \gamma_i), \tag{7.29}$$

where the frequencies g_i ($i = 1, 2$) are the eigenvalues of the matrix $\mathbf{A}$, with e_{ji} the components of the two corresponding eigenvectors, and f_i ($i = 1, 2$) are the eigenvalues of the matrix $\mathbf{B}$, with I_{ji} the components of the corresponding eigenvectors. The phases β_i and γ_i, as well as the amplitudes of the eigenvectors, are determined by the initial conditions. This would correspond to making observations of the osculating eccentricities and inclinations at some time. The solution described by Eqs. (7.28) and (7.29) is the classical *Laplace–Lagrange secular solution* of the secular problem.

With the introduction of the solution to the eigenvalue problem it is easy to confuse the quantities associated with the two bodies and those associated

with the two eigenmodes of the system. In our notation the subscript j always denotes the planet number while the subscript i always denotes the mode number.

It is interesting to note that in our case the characteristic equation for $\mathbf{B}$ is

$$\begin{vmatrix} B_{11} - f & B_{12} \\ B_{21} & B_{22} - f \end{vmatrix} = 0, \tag{7.30}$$

which reduces to

$$f\,[f - (B_{11} + B_{22})] = 0 \tag{7.31}$$

since $B_{11}B_{22} - B_{12}B_{21} = 0$ from the definitions given in Eqs. (7.11) and (7.12). Thus one of the roots of the characteristic equation is $f_1 = 0$ and there is a degeneracy in the problem. This highlights a subtle difference between the $\{h, k\}$ and the $\{p, q\}$ solutions. Whereas an eccentric orbit introduces an asymmetry and a reference line into the problem, a spherical or point-mass central body has no natural reference plane. Physically it is only meaningful to talk about a mutual inclination and hence the choice of a reference plane is arbitrary. For example, it is customary to refer satellite orbits to the equatorial plane of the planet (i.e., the plane perpendicular to its spin vector). However, as we shall see, the introduction of a nonspherical planet adds terms to the diagonal elements of $\mathbf{B}$ and removes the degeneracy problem.

Another point concerning our solution is that it is independent of the mean longitudes because these have been deliberately excluded from the averaged part of the disturbing function. Therefore, although we are able to predict the variations in the eccentricities, inclinations, pericentres, and nodes of the two bodies, we have no information about their positions in space.

The solution given in Eqs. (7.28) and (7.29) implies that the resulting motion of all the masses is stable *for all time*. However, it is important to remember the assumptions under which this result was derived: (i) no mean motion commensurabilities, (ii) $\mathbf{r}_1 < \mathbf{r}_2$, and (iii) the es and Is are small enough that a second-order expansion of the disturbing function is sufficient to describe the motion. But the amplitudes of the eccentricity eigenvectors, for example, could be large enough for the orbits to intersect, violating conditions (ii) and (iii). As we shall see there may be situations where no mean motion commensurabilities exist, but where "small divisor" terms are still important. We have derived a theory that is correct only to the first order in the masses and so it is important to realise that there could be significant contributions from a second-order theory.

7.3 Jupiter and Saturn

We will now apply the theory given above to the case of Jupiter (mass m_1) and Saturn (mass m_2) orbiting the Sun (mass m_c). In 1983 the system had the

following parameters:

$$\begin{aligned} m_1/m_c &= 9.54786 \times 10^{-4}, & m_2/m_c &= 2.85837 \times 10^{-4}, \\ a_1 &= 5.202545 \text{ AU}, & a_2 &= 9.554841 \text{ AU}, \\ n_1 &= 30.3374^\circ \text{y}^{-1}, & n_2 &= 12.1890^\circ \text{y}^{-1}, \\ e_1 &= 0.0474622, & e_2 &= 0.0575481, \\ \varpi_1 &= 13.983865^\circ, & \varpi_2 &= 88.719425^\circ, \\ I_1 &= 1.30667^\circ, & I_2 &= 2.48795^\circ, \\ \Omega_1 &= 100.0381^\circ, & \Omega_2 &= 113.1334^\circ. \end{aligned} \tag{7.32}$$

Since $\alpha = a_1/a_2 = 0.544493$, we can use the definition of Laplace coefficients given in Eqs. (7.13) and (7.14) to get

$$b_{3/2}^{(1)} = 3.17296, \qquad b_{3/2}^{(2)} = 2.07110. \tag{7.33}$$

Using the definitions of the matrix elements given in Eqs. (7.9)–(7.12) we have

$$\mathbf{A} = \begin{pmatrix} +0.00203738 & -0.00132987 \\ -0.00328007 & +0.00502513 \end{pmatrix} {}^\circ\text{y}^{-1} \tag{7.34}$$

and

$$\mathbf{B} = \begin{pmatrix} -0.00203738 & +0.00203738 \\ +0.00502513 & -0.00502513 \end{pmatrix} {}^\circ\text{y}^{-1}. \tag{7.35}$$

We can now find the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ by solving the respective characteristic equations:

$$\begin{vmatrix} A_{11} - g & A_{12} \\ A_{21} & A_{22} - g \end{vmatrix} = g^2 - (A_{11} + A_{22})\, g + (A_{11}A_{22} - A_{21}A_{12}) = 0 \tag{7.36}$$

and

$$\begin{vmatrix} B_{11} - f & B_{12} \\ B_{21} & B_{22} - f \end{vmatrix} = f^2 - (B_{11} + B_{22})\, f + (B_{11}B_{22} - B_{21}B_{12}) = 0. \tag{7.37}$$

The solutions of the resulting quadratic equations are

$$g_1 = 9.63435 \times 10^{-4} \, {}^\circ\text{y}^{-1}, \qquad g_2 = 6.09908 \times 10^{-3} \, {}^\circ\text{y}^{-1} \tag{7.38}$$

and

$$f_1 = 0, \qquad f_2 = -7.06251 \times 10^{-3} \, {}^\circ\text{y}^{-1}. \tag{7.39}$$

The eigenvectors of $\mathbf{A}$ and $\mathbf{B}$ are the four vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}_1$, and $\mathbf{y}_2$ that satisfy the equations

$$\mathbf{A}\mathbf{x}_i = g_i \mathbf{x}_i \qquad \text{and} \qquad \mathbf{B}\mathbf{y}_i = f_i \mathbf{y}_i \qquad (i = 1, 2). \tag{7.40}$$

However, it is clear from these definitions that if $\mathbf{x}_i$ is an eigenvector of the matrix $\mathbf{A}$ then so is $c\mathbf{x}_i$, where c is a constant. Therefore each eigenvector is only determined up to some arbitrary scaling constant. If we let $\bar{e}_{ji}$ and $\bar{I}_{ji}$ denote the

components of these unscaled eigenvectors and let S_i and T_i denote the scaling constant (or magnitude) of each eigenvector, then

$$S_i \bar{e}_{ji} = e_{ji} \qquad \text{and} \qquad T_i \bar{I}_{ji} = I_{ji} \qquad (i = 1, 2). \tag{7.41}$$

The values of $\bar{e}_{ji}$ and $\bar{I}_{ji}$ are obtained by solving four sets of two simultaneous linear equations in two unknowns. The four resulting (unscaled) eigenvectors are

$$\begin{pmatrix} \bar{e}_{11} \\ \bar{e}_{21} \end{pmatrix} = \begin{pmatrix} -0.777991 \\ -0.628275 \end{pmatrix}, \qquad \begin{pmatrix} \bar{e}_{12} \\ \bar{e}_{22} \end{pmatrix} = \begin{pmatrix} 0.332842 \\ -1.01657 \end{pmatrix},$$
$$\begin{pmatrix} \bar{I}_{11} \\ \bar{I}_{21} \end{pmatrix} = \begin{pmatrix} 0.707107 \\ 0.707107 \end{pmatrix}, \qquad \begin{pmatrix} \bar{I}_{12} \\ \bar{I}_{22} \end{pmatrix} = \begin{pmatrix} -0.40797 \\ 1.00624 \end{pmatrix}. \tag{7.42}$$

The scaling factors S_i and T_i are determined from the boundary conditions. At time $t = 0$ we have

$$h_1 = 0.0114692, \quad h_2 = 0.0575337, \quad k_1 = 0.0460556, \quad k_2 = 0.00128611 \tag{7.43}$$

and

$$p_1 = 0.0224566, \quad p_2 = 0.0399314, \quad q_1 = -0.00397510, \quad q_2 = -0.0170597, \tag{7.44}$$

where we have converted the inclinations from degrees to radians. Substituting $t = 0$ in our general solution given in Eqs. (7.28) and (7.29) we get

$$h_j = S_1 \bar{e}_{j1} \sin\beta_1 + S_2 \bar{e}_{j2} \sin\beta_2, \qquad k_j = S_1 \bar{e}_{j1} \cos\beta_1 + S_2 \bar{e}_{j2} \cos\beta_2 \tag{7.45}$$

and

$$p_j = T_1 \bar{I}_{j1} \sin\gamma_1 + T_2 \bar{I}_{j2} \sin\gamma_2, \qquad q_j = T_1 \bar{I}_{j1} \cos\gamma_1 + T_2 \bar{I}_{j2} \cos\gamma_2, \tag{7.46}$$

where the subscript j $(= 1, 2)$ denotes the planet (Jupiter or Saturn). These can be considered as four sets of two simultaneous linear equations in the eight unknowns $S_i \sin\beta_i$, $S_i \cos\beta_i$, $T_i \sin\gamma_i$, and $T_i \cos\gamma_i$ with $(i = 1, 2)$. In our case the solutions are

$$\begin{pmatrix} S_1 \sin\beta_1 \\ S_2 \sin\beta_2 \end{pmatrix} = \begin{pmatrix} -0.0308089 \\ -0.375549 \end{pmatrix}, \qquad \begin{pmatrix} S_1 \cos\beta_1 \\ S_2 \cos\beta_2 \end{pmatrix} = \begin{pmatrix} -0.0472469 \\ 0.027935 \end{pmatrix},$$
$$\begin{pmatrix} T_1 \sin\gamma_1 \\ T_2 \sin\gamma_2 \end{pmatrix} = \begin{pmatrix} 0.0388876 \\ 0.0123566 \end{pmatrix}, \qquad \begin{pmatrix} T_1 \cos\gamma_1 \\ T_2 \cos\gamma_2 \end{pmatrix} = \begin{pmatrix} -0.0109598 \\ -0.00925221 \end{pmatrix}. \tag{7.47}$$

These give

$$\beta_1 = -146.892^\circ, \quad \beta_2 = -53.3565,^\circ \quad \gamma_1 = 105.74^\circ \quad \gamma_2 = 126.825^\circ \tag{7.48}$$

and

$$S_1 = 0.0564044, \quad S_2 = 0.0468053, \quad T_1 = 0.0404025, \quad T_2 = 0.0154366. \tag{7.49}$$

The resulting, scaled eigenvectors are

$$\begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix} = \begin{pmatrix} -0.0438821 \\ -0.0354375 \end{pmatrix}, \quad \begin{pmatrix} e_{12} \\ e_{22} \end{pmatrix} = \begin{pmatrix} 0.0155788 \\ -0.047581 \end{pmatrix},$$
$$\begin{pmatrix} I_{11} \\ I_{21} \end{pmatrix} = \begin{pmatrix} 0.0285689 \\ 0.0285689 \end{pmatrix}, \quad \begin{pmatrix} I_{12} \\ I_{22} \end{pmatrix} = \begin{pmatrix} -0.00629766 \\ 0.015533 \end{pmatrix}, \tag{7.50}$$

where the I_{ji} are expressed in radians.

We have now determined all the constants in Eqs. (7.28) and (7.29). Therefore we can obtain h, k, p, and q for Jupiter and Saturn at any time t. The solution is of the form

$$\begin{aligned} h_j &= e_{j1} \sin(g_1 t + \beta_1) + e_{j2} \sin(g_2 t + \beta_2), \\ k_j &= e_{j1} \cos(g_1 t + \beta_1) + e_{j2} \cos(g_2 t + \beta_2), \\ p_j &= I_{j1} \sin(f_1 t + \gamma_1) + I_{j2} \sin(f_2 t + \gamma_2), \\ q_j &= I_{j1} \cos(f_1 t + \gamma_1) + I_{j2} \cos(f_2 t + \gamma_2), \end{aligned} \tag{7.51}$$

where $j = 1$ for Jupiter and $j = 2$ for Saturn. From these solutions we can derive the orbital elements of the two planets at any time t. For example, the relation $e_j(t) = (h_j^2 + k_j^2)^{1/2}$ is used to calculate the eccentricity of planet j. Using our results we obtain

$$\begin{aligned} e_1(t) &= \sqrt{0.00217 - 0.00137 \cos(93.5^\circ + 0.00514\, t)}, \\ e_2(t) &= \sqrt{0.00352 + 0.00337 \cos(93.5^\circ + 0.00514\, t)}, \end{aligned} \tag{7.52}$$

where the phases are in degrees and the frequencies in degrees per year. This implies a fixed periodicity of $\sim$ 70,100 y in the variation of the eccentricity of each planet. Figure 7.1a shows the evolution of the eccentricities of the two planets over a time span of 200,000 y derived from our secular solution; the periodicity in the variation is clear. The different signs in the magnitude of the cosine imply that a maximum in Jupiter's eccentricity coincides with a minimum in Saturn's eccentricity and vice versa.

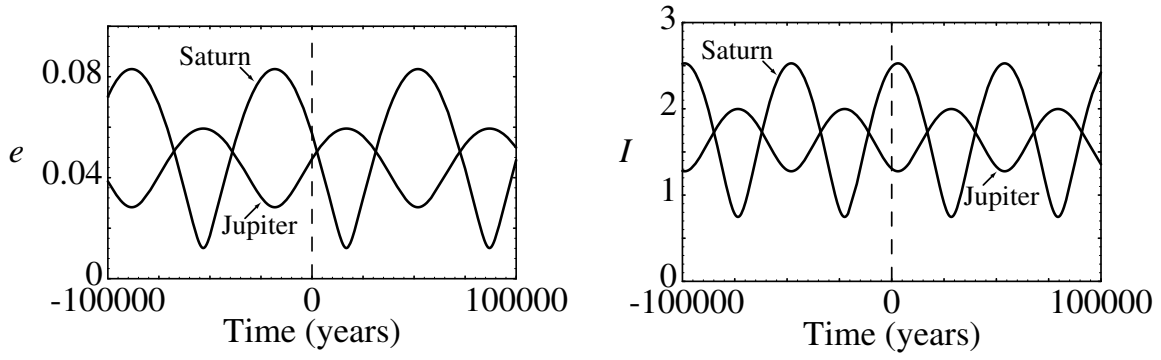

Fig. 7.1. The (a) eccentricities and (b) inclinations of Jupiter and Saturn derived from a secular perturbation theory calculated over a time span of 200,000 y centred on 1983.

Similarly the relation $I_j = (p_j^2 + q_j^2)^{1/2}$ is used to calculate the inclination of planet j at any time t. Our results (in radians) give

$$\begin{aligned} I_1(t) &= \sqrt{0.000856 - 0.00360\cos(21.1^\circ - 0.00706\,t)}, \\ I_2(t) &= \sqrt{0.00106 + 0.000888\cos(21.1^\circ - 0.00706\,t)}. \end{aligned} \tag{7.53}$$

In this case the associated period of the secular variation in each planet is $\sim$51,000 y; since $f_1 = 0$, this period is just $360^\circ/f_2$. The variation for each planet is shown in Fig. 7.1b.

The secular solution that we have derived for Jupiter and Saturn is only an approximation to the actual variations in their orbital elements. In reality the perturbations from the planets Uranus and Neptune exert considerable influence on their orbits. A further complication is that the orbits of Jupiter and Saturn are close to a 5:2 commensurability. This introduces additional perturbations on timescales that are shorter than those associated with the secular variation.

7.4 Free and Forced Elements

We have shown that under certain conditions we can construct a secular solution to the motion of two orbiting bodies moving under their mutual gravitational effects; at any time we can obtain the eccentricities, longitudes of pericentre, inclinations, and longitudes of ascending node of both bodies. We can make use of this solution to study the motion of an additional body, of negligible mass, moving under the influence of the central body and perturbed by the other two bodies.

Following the example given in Sect. 7.2 for the secular theory for two bodies, the disturbing function $\mathcal{R}$ for a test particle with orbital elements a, n, e, I, ϖ, and Ω is given by

$$\mathcal{R} = na^2\Bigg[\frac{1}{2}Ae^2 + \frac{1}{2}BI^2 + \sum_{j=1}^{2} A_j e e_j \cos(\varpi - \varpi_j) + \sum_{j=1}^{2} B_j I I_j \cos(\Omega - \Omega_j)\Bigg], \tag{7.54}$$

where

$$A = +n\frac{1}{4}\sum_{j=1}^{2}\frac{m_j}{m_c}\alpha_j\bar{\alpha}_j b_{3/2}^{(1)}(\alpha_j), \tag{7.55}$$

$$A_j = -n\frac{1}{4}\frac{m_j}{m_c}\alpha_j\bar{\alpha}_j b_{3/2}^{(2)}(\alpha_j), \tag{7.56}$$

$$B = -n\frac{1}{4}\sum_{j=1}^{2}\frac{m_j}{m_c}\alpha_j\bar{\alpha}_j b_{3/2}^{(1)}(\alpha_j), \tag{7.57}$$

$$B_j = +n\frac{1}{4}\frac{m_j}{m_c}\alpha_j\bar{\alpha}_j b_{3/2}^{(1)}(\alpha_j) \tag{7.58}$$

and

$$\alpha_j = \begin{cases} a_j/a & \text{if} \quad a_j < a, \\ a/a_j & \text{if} \quad a_j > a, \end{cases} \tag{7.59}$$

$$\bar{\alpha}_j = \begin{cases} 1 & \text{if} \quad a_j < a, \\ a/a_j & \text{if} \quad a_j > a. \end{cases} \tag{7.60}$$

If we now transform to a new set of variables h, k, p, and q for the test particle and h_j, k_j, p_j, and q_j $(j = 1, 2)$ for each perturbing body, where

$$h = e \sin \varpi, \qquad k = e \cos \varpi \tag{7.61}$$

and

$$p = I \sin \Omega, \qquad q = I \cos \Omega \tag{7.62}$$

and the other elements are already defined in Eqs. (7.18) and (7.19), we have

$$\mathcal{R} = na^2 \Bigg[\frac{1}{2} A(h^2 + k^2) + \frac{1}{2} B(p^2 + q^2) + \sum_{j=1}^{2} A_j (h h_j + k k_j) + \sum_{j=1}^{2} B_j (p p_j + q q_j) \Bigg]. \tag{7.63}$$

The equations of motion are

$$\dot{h} = +\frac{1}{na^2} \frac{\partial \mathcal{R}}{\partial k}, \qquad \dot{k} = -\frac{1}{na^2} \frac{\partial \mathcal{R}}{\partial h}, \tag{7.64}$$

$$\dot{p} = +\frac{1}{na^2} \frac{\partial \mathcal{R}}{\partial q}, \qquad \dot{q}_j = -\frac{1}{na^2} \frac{\partial \mathcal{R}}{\partial p}. \tag{7.65}$$

Substituting for $\mathcal{R}$ from Eq. (7.63) we can write the equations of motion as

$$\dot{h} = +Ak + \sum_{j=1}^{2} A_j k_j, \qquad \dot{k} = -Ah - \sum_{j=1}^{2} A_j h_j, \tag{7.66}$$

$$\dot{p} = +Bq + \sum_{j=1}^{2} B_j q_j, \qquad \dot{q} = -Bp - \sum_{j=1}^{2} B_j p_j, \tag{7.67}$$

where the values of h_j, k_j, p_j, and q_j are derived from the secular solution given in Eqs. (7.28) and (7.29). Substituting from these equations we get

$$\dot{h} = +Ak + \sum_{j=1}^{2} A_j \sum_{i=1}^{2} e_{ji} \cos(g_i t + \beta_i), \tag{7.68}$$

$$\dot{k} = -Ah - \sum_{j=1}^{2} A_j \sum_{i=1}^{2} e_{ji} \sin(g_i t + \beta_i), \tag{7.69}$$

$$\dot{p} = +Bq + \sum_{j=1}^{2} B_j \sum_{i=1}^{2} I_{ji} \cos(f_i t + \gamma_i), \tag{7.70}$$

$$\dot{q} = -Bp - \sum_{j=1}^{2} B_j \sum_{i=1}^{2} I_{ji} \sin(f_i t + \gamma_i). \tag{7.71}$$

By taking another time derivative of each equation and using Eq. (7.26) again we have

$$\ddot{h} = -A^2 h - \sum_{i=1}^{2} \nu_i (A + g_i) \sin(g_i t + \beta_i), \tag{7.72}$$

$$\ddot{k} = -A^2 k - \sum_{i=1}^{2} \nu_i (A + g_i) \cos(g_i t + \beta_i), \tag{7.73}$$

$$\ddot{p} = -B^2 p - \sum_{i=1}^{2} \mu_i (B + f_i) \sin(f_i t + \gamma_i), \tag{7.74}$$

$$\ddot{q} = -B^2 q - \sum_{i=1}^{2} \mu_i (B + f_i) \cos(f_i t + \gamma_i), \tag{7.75}$$

where

$$\nu_i = \sum_{j=1}^{2} A_j e_{ji} \qquad \text{and} \qquad \mu_i = \sum_{j=1}^{2} B_j I_{ji}. \tag{7.76}$$

The solutions to the uncoupled differential equations in Eqs. (7.72)–(7.75) are

$$\begin{aligned} h &= e_{\text{free}} \sin(At + \beta) + h_0(t), & k &= e_{\text{free}} \cos(At + \beta) + k_0(t), \\ p &= I_{\text{free}} \sin(Bt + \gamma) + p_0(t), & q &= I_{\text{free}} \cos(Bt + \gamma) + q_0(t), \end{aligned} \tag{7.77}$$

where e_{free}, I_{free}, β, and γ are constants determined from the boundary conditions and

$$h_0(t) = -\sum_{i=1}^{2} \frac{\nu_i}{A - g_i} \sin(g_i t + \beta_i), \tag{7.78}$$

$$k_0(t) = -\sum_{i=1}^{2} \frac{\nu_i}{A - g_i} \cos(g_i t + \beta_i), \tag{7.79}$$

$$p_0(t) = -\sum_{i=1}^{2} \frac{\mu_i}{B - f_i} \sin(f_i t + \gamma_i), \tag{7.80}$$

$$q_0(t) = -\sum_{i=1}^{2} \frac{\mu_i}{B - f_i} \cos(f_i t + \gamma_i). \tag{7.81}$$

Note that h_0, k_0, p_0, and q_0 are only functions of the (constant) semi-major axis of the particle and do not involve any of its other orbital elements. However,

since the values of h_0, k_0, p_0, and q_0 also depend on the secular solution for the two perturbing bodies, they will vary with time.

If we define the quantities

$$e_{\text{forced}} = \sqrt{h_0^2 + k_0^2}, \qquad I_{\text{forced}} = \sqrt{p_0^2 + q_0^2} \tag{7.82}$$

then the solutions given in Eq. (7.77) have a simple geometrical interpretation (Figs. 7.2 and 7.3). In the case of the h–k solution, the values of k and h for the particle define a point in the k–h plane. The vector from the origin to this point has a length e and it makes an angle ϖ with the k axis. In the light of our solution given above, this vector can also be thought of as the vector sum of two other vectors: The first goes from the origin to the point (k_0, h_0); it has a length e_{forced} and makes an angle ϖ_{forced} with the k axis. The second goes from (k_0, h_0) to the point (k, h); it has a length e_{free} and makes an angle $\varpi_{\text{free}} = At + \beta$ with the k axis. This implies that the particle's motion can be thought of as motion around a circle with centre (k_0, h_0) at a constant rate A while this point itself moves in some complicated path determined by the secular solution for the two perturbing bodies. This is illustrated in Fig. 7.2. The quantities e_{forced} and ϖ_{forced} are derived from h_0 and k_0 and they are called the *forced eccentricity* and *forced longitude of pericentre* of the particle. Their values are determined solely by the semi-major axis of the particle and the secular solution for the two perturbing bodies. In contrast, e_{free} and ϖ_{free}, the *free eccentricity* and *free longitude of pericentre* of the particle, are derived from the boundary conditions and denote fundamental orbital parameters of the particle. These quantities are also referred to as the *proper eccentricity* and *proper longitude of pericentre* of the particle's orbit.

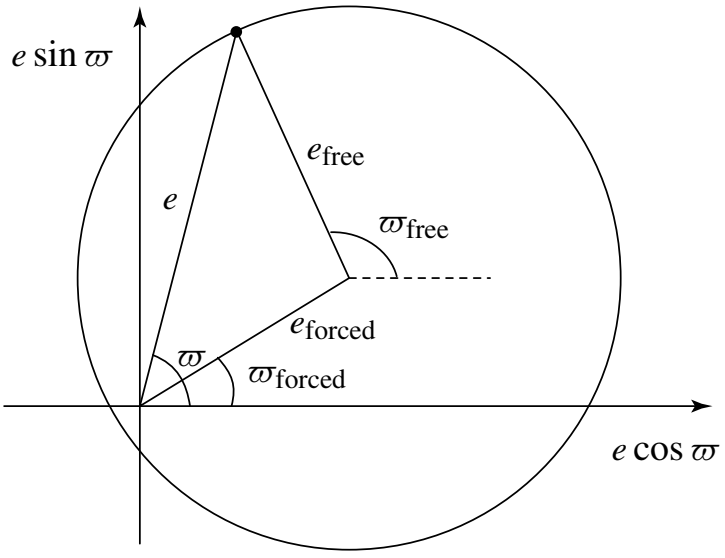

Fig. 7.2. The geometrical relationship among the osculating, free, and forced eccentricities and longitudes of pericentre for the case $e_{\text{free}} > e_{\text{forced}}$.

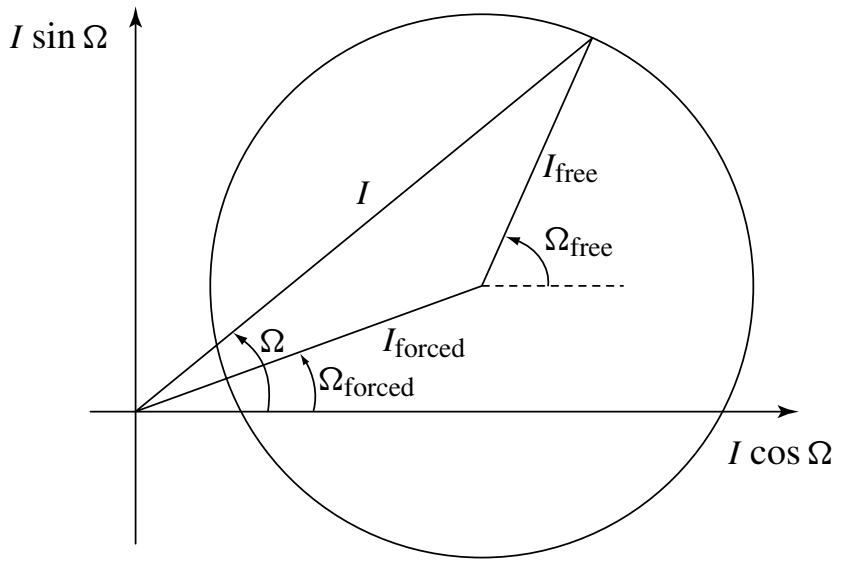

Fig. 7.3. The geometrical relationship among the osculating, free, and forced inclinations and longitudes of ascending node for the case $I_{\rm free} < I_{\rm forced}$.

It is important to note that the circle shown in Fig. 7.2 need not encompass the origin. If $e_{\rm free}$ is small enough or the value of $e_{\rm forced}$ for the particle's semi-major axis is large enough, the motion of the particle around the circle might be such that ϖ or Ω (the osculating longitudes of pericentre or ascending node) vary over some fixed range of angles.

The p–q solution is shown in Fig. 7.3, where $I_{\rm forced}$, $\Omega_{\rm forced}$, $I_{\rm free}$, and $\Omega_{\rm free}$ denote the forced and free inclinations and nodes of the particle. Here we illustrate a situation where the forced inclination is larger than the free one, such that the circle does not enclose the origin. As stated above, this is an outcome of the boundary conditions.

Because the expressions for $e_{\rm forced}$ and $I_{\rm forced}$ given in Eq. (7.82) depend on the definitions of h_0, k_0, p_0, and q_0 given in Eqs. (7.78)–(7.81), it is clear that potentially large values of $e_{\rm forced}$ or $I_{\rm forced}$ can arise if either of the conditions $A - g_i \approx 0$ or $B - f_i \approx 0$ is satisfied. The g_i and f_i are the eigenfrequencies of the system of two interacting bodies whereas, from Eqs. (7.55) and (7.57), the quantities A and B are functions of the semi-major axis of the test particle. This implies that at certain locations in semi-major axis there will be singularities in the forced eccentricities or inclinations. We will consider a specific example of this in Sect. 7.5.

Another important point concerns the limiting values of $e_{\rm forced}$ and $I_{\rm forced}$ as the orbit of either of the two perturbers is approached. Since the A, A_j, B, and B_j in Eqs. (7.55)–(7.58) as well as the definitions of the ν_i and μ_i in Eq. (7.76) involve the Laplace coefficients $b^{(1)}_{3/2}(\alpha_j)$ or $b^{(2)}_{3/2}(\alpha_j)$, which all approach infinity as $\alpha_j \to 1$, it is not obvious that there are finite limiting values for $e_{\rm forced}$ and $I_{\rm forced}$ at the orbits of the perturbers. Let us assume that we are considering the behaviour of $e_{\rm forced}$ in the vicinity of a perturbing body denoted by subscript

$j = l$. In this case

$$b_{3/2}^{(1)}(\alpha_l) \to \infty \qquad \text{and} \qquad b_{3/2}^{(2)}(\alpha_l) \to \infty \qquad \text{as} \qquad \alpha_l \to 1\,. \tag{7.83}$$

At the orbit of the perturber $A \gg g_i$ for $i = 1, 2$ and $A_l \gg A_i$, where $i \neq l$. Therefore

$$A - g_i \approx A \approx \frac{1}{4} n \frac{m_l}{m_\mathrm{c}} \alpha_l \bar{\alpha}_l b_{3/2}^{(1)}(\alpha_l) \tag{7.84}$$

and

$$\nu_i = \sum_{j=1}^{2} A_j e_{ji} \approx A_l e_{li}. \tag{7.85}$$

This implies that

$$h_0(t) \approx -\sum_{i=1}^{2} \frac{A_l e_{li}}{A} \sin(g_i t + \beta_i) = +\frac{b_{3/2}^{(2)}(\alpha_l)}{b_{3/2}^{(1)}(\alpha_l)} \sum_{i=1}^{2} e_{li} \sin(g_i t + \beta_i). \tag{7.86}$$

Since the definition of the Laplace coefficient given in Eqs. (6.67) and (6.68) can also be written as

$$\frac{1}{2} b_s^{(j)}(\alpha) = \frac{s(s+1)\ldots(s+j-1)}{j!} \alpha^j F(s, s+j, j+1; \alpha^2), \tag{7.87}$$

where $F(a, b, c; d)$ is the standard hypergeometric function, we have

$$\lim_{\alpha_l \to 1} \frac{b_{3/2}^{(2)}(\alpha_l)}{b_{3/2}^{(1)}(\alpha_l)} = \lim_{\alpha_l \to 1} \left[\frac{5}{4} \alpha_l \frac{F(\frac{3}{2}, \frac{7}{2}, 3, \alpha_l^2)}{F(\frac{3}{2}, \frac{5}{2}, 2, \alpha_l^2)} \right] = 1 \tag{7.88}$$

from the properties of hypergeometric series and their relationship with elliptical integrals. Therefore

$$\lim_{\alpha_l \to 1} h_0(t) = \sum_{i=1}^{2} e_{li} \sin(g_i t + \beta_i) = h_l. \tag{7.89}$$

Similarly

$$\lim_{\alpha_l \to 1} k_0(t) = \sum_{i=1}^{2} e_{li} \cos(g_i t + \beta_i) = k_l. \tag{7.90}$$

Therefore, as the orbit of perturbing body l is approached,

$$e_\mathrm{forced} = \sqrt{h_0^2 + k_0^2} \to \sqrt{h_l^2 + k_l^2} = e_l. \tag{7.91}$$

This implies that the forced values of the eccentricity and longitude of pericentre at the orbit of the perturber are equal to the equivalent osculating values of these elements for the perturber. A similar result for the forced values of the inclination and longitude of ascending node can be shown using the same method as given above. However, we have to be careful not to make too many generalisations

about the nature of the particle's orbit in the vicinity of the orbit of a perturber. We have already seen in Fig. 3.30 that particles near the L_1 and L_2 points acquire an eccentricity from their encounter with the satellite, despite the fact that the satellite is moving in a circular orbit. Therefore our result concerning the forced eccentricity and inclination is really only valid for particle orbits far from L_1 and L_2 with low values of e and I.

7.10 Hirayama Families and the IRAS Dust Bands

The analysis given in Sect. 7.4 and Sect. 7.9 implies that the osculating eccentricity of a small mass object, such as an asteroid or a dust particle, moving under the gravitational effects of the planets can be thought of as having two components: (i) a free or proper eccentricity, which reflects the "inherent" eccentricity of the object, and (ii) a forced eccentricity, which is a result of its current semi-major axis and the relative positions of its perturbers. The same arguments apply to the inclination of the object. Therefore the free or proper elements provide information on the body's natural elements rather than those that are matters of its circumstances.

Hirayama (1918) derived proper elements for the limited number of asteroids known at the time. He showed that some asteroids tended to cluster in groups in (a, e) and (a, I) space and that the clustering was more pronounced when plotted using free rather than osculating elements. He proposed that each cluster, or *family*, represented objects that had a common dynamical origin and were the remnants of the breakup of a parent body. It is customary to name an identified family after its largest member. Today, soon after a valid orbit has been determined, an asteroid's proper elements are calculated and family associations are examined. On this basis almost half of all asteroids are thought to be family members.

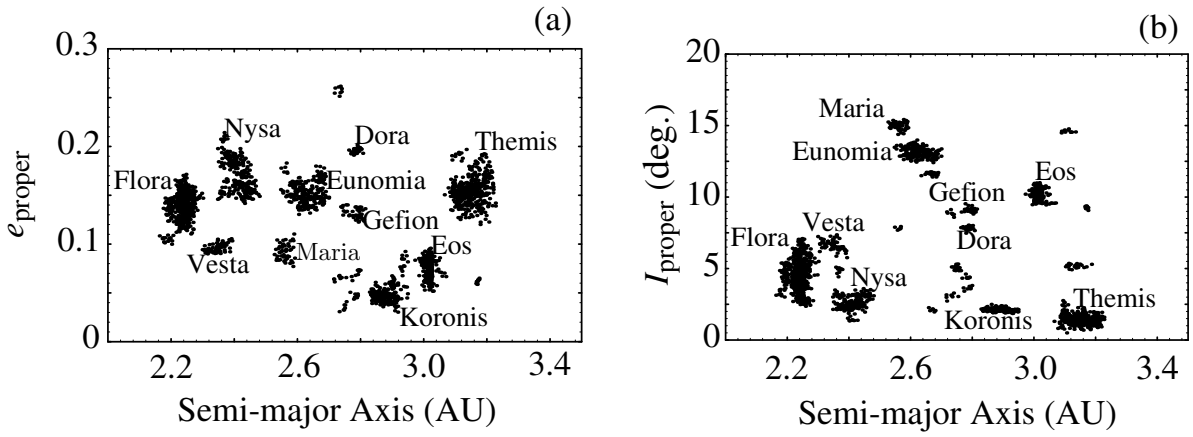

Fig. 7.13. A plot of (a) the proper (or free) eccentricity and (b) the proper inclination of asteroids in the range $2.0 \leq a \leq 3.5$ AU that have been identified as family members.

Figure 7.13a shows the proper eccentricity as a function of semi-major axis for those asteroids that have been identified as belonging to families. The equivalent diagram for proper inclination is shown in Fig. 7.13b. There are a number of obvious clusterings of asteroids and it is important to realise that these are groupings in three-dimensional (a, e, I) space.

We have already shown that the variation of the orbital elements can be thought of as motion around a circle with a centre determined by the forced component (see Figs. 7.2 and 7.3). In Fig. 7.14 we illustrate this by plotting osculating values of $k = e\cos\varpi$ versus $h = e\sin\varpi$ and $q = I\cos\Omega$ versus $p = I\sin\Omega$ for the

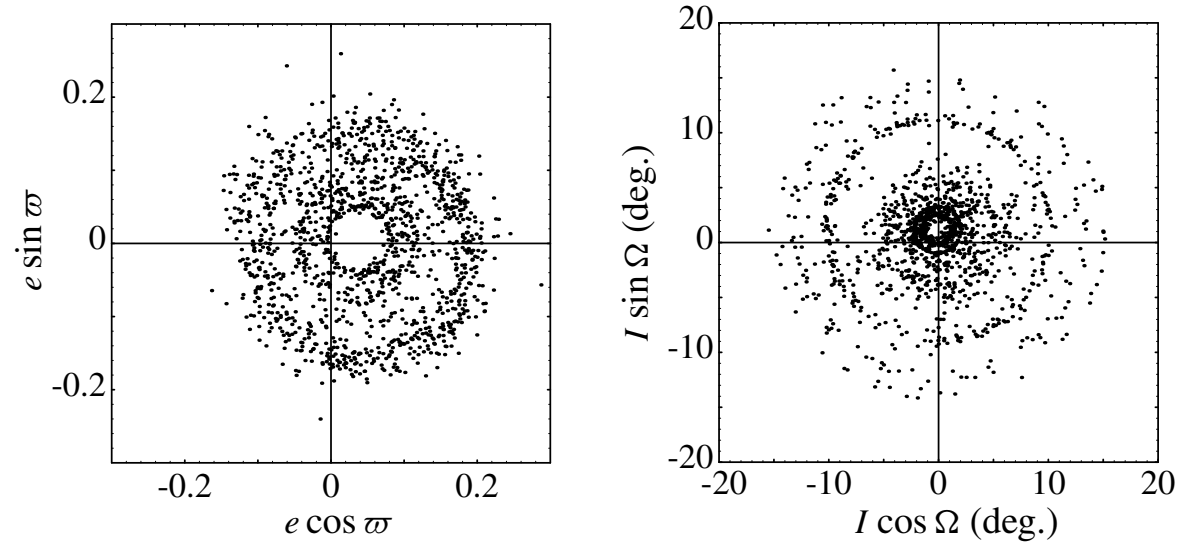

Fig. 7.14. (a) A plot of the osculating values of $k = e\cos\varpi$ and $h = e\sin\varpi$ for all the asteroids shown in Fig. 7.13a. A number of "rings" are clearly visible. (b) A plot of the osculating values of $q = I\cos\Omega$ and $p = I\sin\Omega$ for all the asteroids shown in Fig. 7.13b. A number of "rings" are clearly visible, although most have a small radius (i.e., a low inclination).

Table 7.4. The proper and forced orbital elements of the Koronis, Eos, and Themis asteroid families.

Family	a (AU)	$e_{\rm proper}(^\circ)$	$I_{\rm proper}(^\circ)$	$e_{\rm forced}$	$\varpi_{\rm forced}(^\circ)$	$I_{\rm forced}(^\circ)$	$\Omega_{\rm forced}(^\circ)$
Koronis	2.875	0.049	2.12	0.037	6.2	1.16	96.1
Eos	3.015	0.071	10.20	0.037	7.6	1.19	97.1
Themis	3.136	0.152	1.42	0.038	8.7	1.22	97.8

family members in our sample. In each case the presence of "circles" of asteroids is clearly visible. This implies that members of a given asteroid family do indeed have a common forced eccentricity and inclination (determined by their semi-major axis) and a common free eccentricity and inclination with randomised free pericentres and nodes. Note that, as predicted from secular theory, the centres of these circles are not identical.

Three of the largest clusterings are associated with the Koronis, Eos, and Themis families, all of which are thought to be the fragments of larger bodies that suffered catastrophic collisions at least 10^7 years ago. The proper and forced orbital elements of these families are listed in Table 7.4. The data used in this table are taken from Dermott *et al.* (1985). If these families were formed by the collisional breakup of a larger body then such an event would have produced a large amount of asteroidal dust. Dramatic evidence for the collisional theory for the origin of the Hirayama families came with the discovery by the Infrared Astronomical Satellite (IRAS) of dust bands in the solar system (Low et al. 1984, Neugebauer et al. 1984).

IRAS carried out an all-sky survey at wavelengths of 12, 25, 60, and 100 μm. Surveys at these wavelengths are particularly adept at detecting infrared radiation from dust in the solar system. The background flux detected by IRAS at 25μm is shown in Fig. 7.15a. In this plot the IRAS dust bands are barely detectable as small "bumps" in the distribution close to ecliptic latitudes of 0° and $\pm 10^\circ$. However, if the background component of the curve is removed and the smoothed residuals are plotted (Fig. 7.15b), then the bands are clearly visible – there is a central band that appears to be split and two side bands at $\pm 10^\circ$. But how do we know that these bands originate from the collisions that formed the main families? The answer lies in an understanding of secular perturbations and orbital geometry.

We have already noted that a test particle's forced elements are a function of its semi-major axis alone. Figure 7.16 shows the values of the forced elements in the region of the asteroid belt according to the theory of Brouwer & van Woerkom (1950). Fragments from the breakup of an asteroid would have approximately the same free eccentricity and free inclination but would quickly have acquired

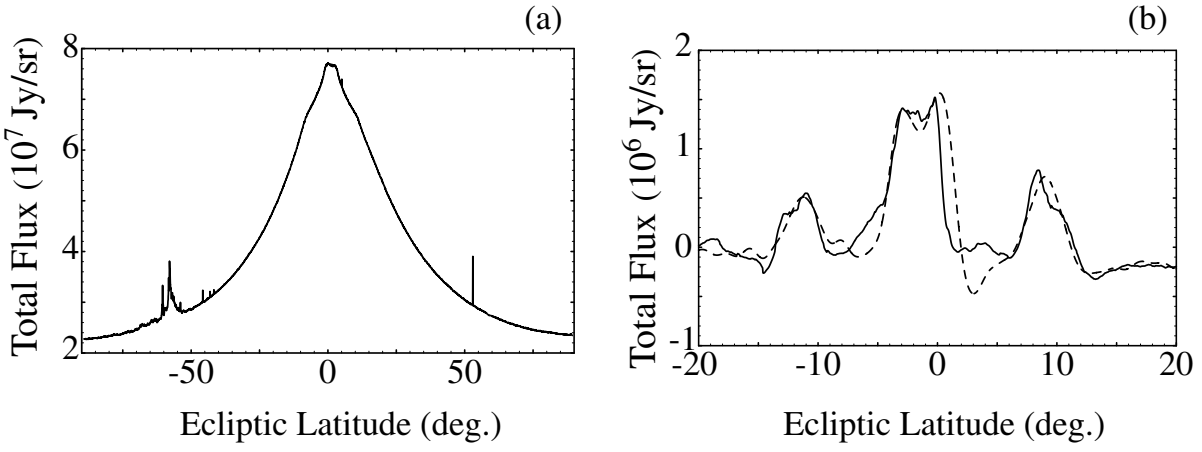

Fig. 7.15. (a) The IRAS background infrared flux at a wavelength of 25μm. The spikes close to ecliptic latitudes -60° and $+50^\circ$ are caused by residual contributions from the galactic plane. (b) The smoothed, residual data showing the flux associated with the dust bands (solid line) together with the model data (dashed line) obtained by assuming that the signal comes from dust in the same distribution of orbits as the Themis, Eos, and Koronis families of asteroids.

randomised free pericentres and free nodes. Having the same forced inclination means that all the fragments precess about a common mean plane determined by the forced inclination and forced node. Because the vertical component of the their motion with respect to this plane is simple harmonic in form, the asteroids will spend most of their time at the extremes of their motion, giving rise to a bunching at these locations. The result is that, viewed from the Sun, the asteroids will give the appearance of lying in two bands separated by $2I_{\rm forced}$ in latitude (see Fig. 7.17).

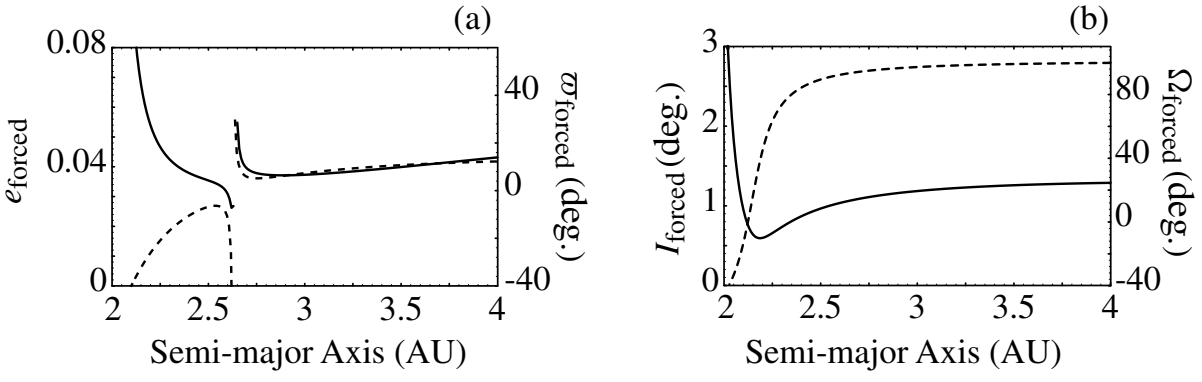

Fig. 7.16. Forced orbital elements as a function of semi-major axis using the secular theory of Brouwer & van Woerkom (1950). (a) The forced eccentricity (solid curve, left-hand axis) and forced longitude of perihelion (dashed curve, right-hand axis). (b) The forced inclination (solid curve, left-hand axis) and forced longitude of ascending node (dashed curve, right-hand axis).

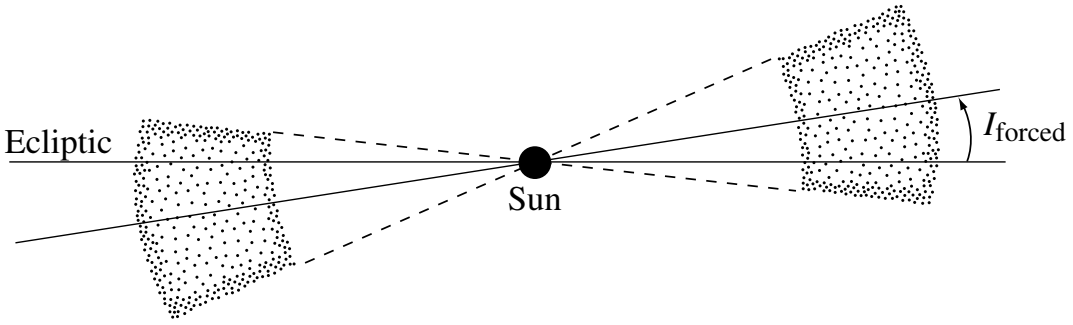

Fig. 7.17. A vertical cross-sectional view of a distribution of asteroids with the same free and forced inclination but randomised free nodes. The radial extent of the cross section is due to the eccentricity of the asteroids.

However, there is also an effect due to the eccentricity. This is shown in Fig. 7.18. Orbits with the same forced and free eccentricities and the same forced pericentre but randomised free pericentres generate an orbital distribution that is symmetrical with respect to a point C that is not the Sun (denoted by the point S in Fig. 7.18). If we envisage asteroidal dust moving in these orbits then the combined effect of the inclination and eccentricity is to produce a cloud of

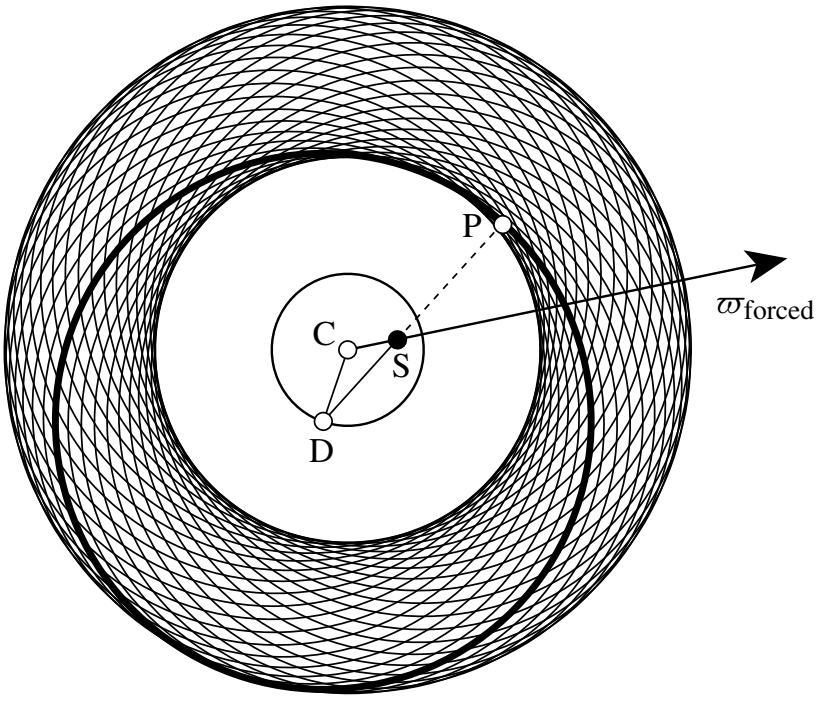

Fig. 7.18. The distribution of eccentric orbits having the same semi-major axis a, the same forced eccentricity e_{forced} and proper eccentricity e_{proper}, the same forced pericentre ϖ_{forced}, but randomised proper longitudes of pericentre. S denotes the position of the Sun and C is the centre of symmetry. An individual orbit with pericentre at P and centre at D is highlighted. In the diagram the lines CS, DS, and DC have lengths ae_{forced}, ae, and ae_{proper} respectively, where e is the eccentricity of the orbit. Note that the centre of symmetry is not the Sun.

material that is not symmetrically placed with respect to the Sun and the ecliptic. The added complication of the spacecraft's geocentric view implies that the appearance of the bands should change depending on the time of year (i.e., the position of Earth in its orbit).

Observations of the varying appearance of the dust bands and the implied forced orbital elements have been used by Dermott et al. (1992) to model the distribution of material. Using sources of dust associated with the Themis, Koronis, Eos, Nysa, Dora, and Gefion families Dermott et al. (1992) used an iterative procedure to produce a model profile (dashed line in Fig. 7.15b) that gives excellent agreement with the observations. The implied association of the dust with the collisions that formed the major Hirayama families appears to be confirmed.

7.11 Secular Resonance

In our study of Brouwer & van Woerkom's (1950) secular perturbation theory for the solar system (Sect. 7.9) we noted that there are problems in calculating the forced elements of test particles at those locations in semi-major axis where either of the proper precession rates (denoted by A or B) of the particle equals one of the eigenfrequencies of the system. In the case of the asteroid belt we noted three such locations: two near 2 AU and one near 2.6 AU. The latter explains the obvious singularity in the calculation of the forced eccentricity and forced longitude of pericentre shown in Fig. 7.16a. This is an example of *secular resonance.*

A resonance arises when two periods or frequencies are in a simple numerical ratio. We have already seen examples of this in Sect. 5.4 where the frequencies in question were the orbital and spin rates of a satellite (or planet). In the case of secular resonance the relevant frequencies are the rates of change of the proper longitude of pericentre ($A = \dot{\varpi}_{\text{proper}}$) or proper longitude of ascending node ($B = \dot{\Omega}_{\text{proper}}$) of the test body (usually an asteroid) and one of the eigenfrequencies of the system of perturbing bodies. Unfortunately the techniques required to analyse these resonances are not as simple as in the spin–orbit case. The basic secular theory used throughout this chapter is based on an expansion of the disturbing function to second degree in the eccentricities and inclinations and use of Lagrange's equations in their lowest order form. This produces a system in which the (e, ϖ) solution is completely decoupled from the (I, Ω) solution. Although this is sufficient to allow us to suggest where secular resonances might occur, a more complete theory requires that higher order terms be taken into account. Furthermore, it is also necessary to take into account terms of the second order in the masses, whereas we have only considered a first-order theory. This makes the mathematics more difficult and introduces a coupling between the eccentricity and inclination terms. For further details see

the review articles by Knežević & Milani (1994) and Froeschlé & Morbidelli (1994).

Williams (1969) derived a semianalytical secular theory without using an expansion of the disturbing function. His calculation of proper elements and subsequent identification of asteroid families (Williams 1979) were fundamental advances in modern studies of asteroid dynamics. Because of coupling the location of secular resonances actually correspond to surfaces in (a, e, I) space rather than being centred on a single semi-major axis. The location of these surfaces for the asteroid belt has been calculated by Williams (1969) and Williams & Faulkner (1981). They studied the secular resonances where the frequencies $A - g_5$, $A - g_6$, and $B - f_6$ are all approximately zero. These are referred to as the *linear secular resonances* and, as we have noted, their existence is suggested by the secular perturbation theory developed in this chapter. These are also known as the ν_5, ν_6, and ν_{16} secular resonances, where the suffix denotes the ith value of the eigenfrequency involved ($\nu_1 = g_1, \ldots, \nu_{10} = g_{10}, \nu_{11} = f_1, \ldots, \nu_{18} = f_8$).

Figure 7.19 shows the location in proper semi-major axis–proper inclination space of the linear secular resonances in the asteroid belt, calculated for $e_{\text{proper}} = 0.1$ (after Milani & Knežević 1990). These are superimposed on the

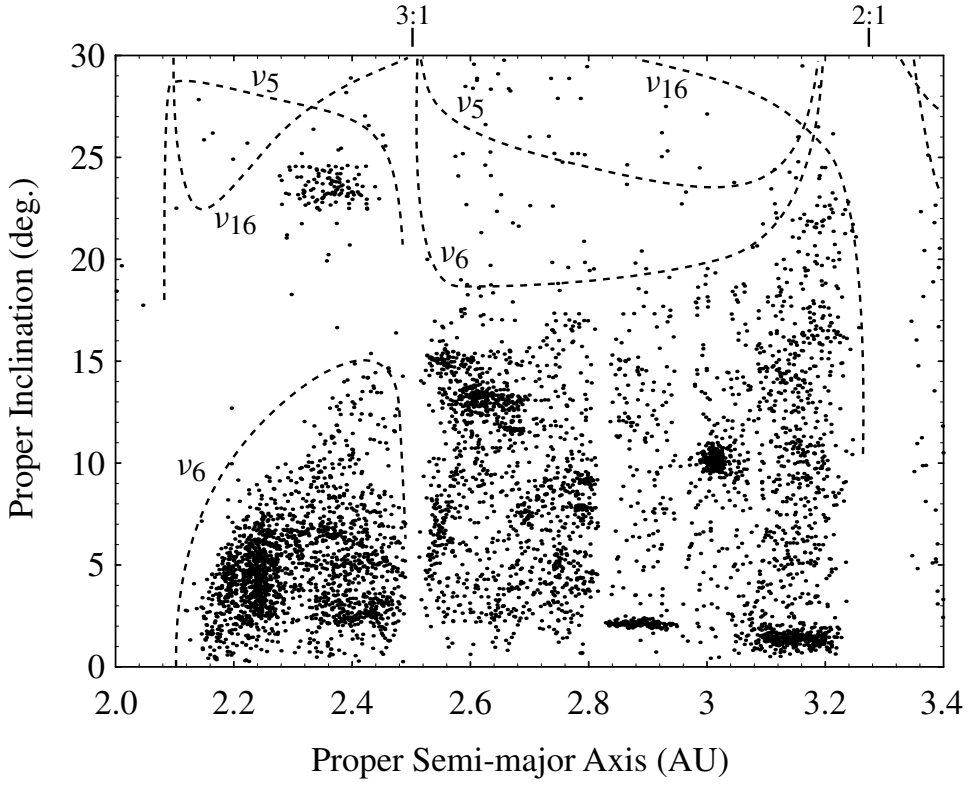

Fig. 7.19. The location of the important ν_5, ν_6, and ν_{16} linear secular resonances (calculated using $e_{\text{proper}} = 0.1$) and the numbered asteroids' I_{proper} as a function of proper semi-major axis.

actual distribution of the proper elements for asteroids in the main belt. For $\sin I_{\text{proper}} < 0.3$ the elements were calculated by Milani & Knežević (1990), using a theory developed by Yuasa (1973), while the elements of Lemaître & Morbidelli (1994) were used for larger inclinations. Note the singularities in Fig. 7.19 close to 2.5 and 3.3 AU. These are the locations of the 3:1 and 2:1 mean motion resonances (see Chapter 8) where the orbital period of the asteroid is a simple fraction of Jupiter's period. At these locations it is necessary to use a secular theory that incorporates the effect of mean motion resonance. The additional singularity close to 2 AU has already been mentioned (see above and Fig. 7.12a). The distribution of asteroids is clearly nonrandom, partly due to the existence of the Kirkwood gaps at a number of resonant locations (see Fig. 1.7 and Sect. 9.8). However, it is also clear that the inner edge of the main belt in a–I space is correlated with the location of the ν_6 secular resonance for $I < 15^\circ$. At higher inclinations there appears to be a group isolated by the three secular resonances and the 3:1 Kirkwood gap.

Other secular resonances are possible, subject to a d'Alembert-like relationship on the permitted combinations of frequencies. These are the *nonlinear secular resonances* and all involve higher powers of eccentricity and/or inclination in the equations of motion. Nine of them ($A + B - g_5 - f_6$, $A + B - g_6 - f_6$, $A + B - g_5 - f_7$, $A - 2g_6 + g_5$, $A - 2g_6 + g_7$, $A - 3g_6 + 2g_5$, $B - f_6 - g_5 + g_6$, $2A + B - 2g_6 - f_6$, and $3A + B - 3g_6 - f_6$) give rise to important secular resonances in the asteroid belt (Knežević & Milani 1994).

Another form of secular resonance exists for small objects on highly inclined orbits, although it does not involve any of the eigenfrequencies of the system. A *Kozai resonance* occurs when $\dot{\omega} = 0$, where ω is the argument of pericentre. Because $\varpi = \omega + \Omega$, the resonance condition reduces to $A = B$. Note that for low-eccentricity, low-inclination orbits in the absence of oblateness A and B are equal in magnitude and opposite in sign. The circumstances under which $A = B$ only occur for highly inclined orbits. It can be shown that the problem of a massless body moving under the gravitational effect of planets moving in coplanar, circular orbits reduces to a system of one degree of freedom, provided there are no resonances between the mean motions (see Chapter 8). Kozai (1962) showed that an asteroid perturbed by Jupiter moving in a circular orbit would have no secular change in its semi-major axis but its eccentricity and inclination could undergo changes such that the quantity

$$H_{\text{K}} = \sqrt{1 - e^2} \cos I \tag{7.151}$$

always remains constant. Note that for constant semi-major axis this is just another way of stating that the third Delaunay momentum, H, is constant (see Eq. (2.176)). This is also related to the Tisserand relation discussed in Sect. 3.4.

A consequence of this constant is that the eccentricity and inclination of a small object's orbit are coupled such that e is a maximum when I is a minimum and *vice versa*.

Kozai's theory was extended by Michel & Thomas (1995) to include the four giant planets. The theory shows that for low inclinations it is possible for ω to librate about stable points at $\omega = 0^\circ$ and $\omega = 180^\circ$. For inclinations greater that $\sim 30^\circ$ these points become unstable and new stable equilibrium points appear at $\omega = 90^\circ$ and $\omega = 270^\circ$. Thomas & Morbidelli (1996) have shown that the Kozai resonance can only affect orbits with large e and I. They also confirmed the conclusion of Bailey et al. (1992) that the Kozai resonance is the mechanism by which some long-period comets can become sungrazing.

7.12 Higher Order Secular Theory

In this chapter we have used the secular theory of Brouwer & van Woerkom (1950) as our adopted theory for the long-term variation of the planetary orbits. This is referred to a linear theory and for the most part it follows the methods of Sect. 7.7 in that it is only to first order in the masses and is based on an expansion of the disturbing function to second degree in the eccentricities and inclinations. Using a treatment due to Hill (1897), Brouwer & van Woerkom made some attempt to incorporate a higher order theory to account for the Jupiter–Saturn interactions. These interactions give rise to the g_9 and g_{10} eigenfrequencies and their associated eigenvectors listed in Tables 7.1–7.3.

Incorporating higher degree terms in the orbital elements, Bretagnon (1974) developed a secular theory for all the planets except planets to second order in the masses. A later version incorporated the effects of relativity and lunar perturbations (Bretagnon 1982). Following the methods of Duriez (1979), a new secular theory was devised by Laskar (1985, 1986a). This included the same perturbations as Bretagnon's later theory, but with the addition of even higher order terms in the eccentricities and inclinations. The resulting theory was numerically integrated over a time span of 30 My and Fourier analysed (Laskar 1988). Around the same time a number of researchers undertook extensive numerical integrations of the outer solar system in studies of long-term stability of planetary orbits (Kinoshita & Nakai 1984, Applegate et al. 1986, Carpino et al. 1987). There have also been integrations of the orbits of the inner planets. For example, Quinn et al. (1991) investigated the Earth's orbit over a timescale of 3 My. Their results, subsequently extended to 6 My, were compared with the semianalytical secular theory of Laskar (1989) and were found to be in good agreement (Laskar et al. 1992). All of these studies provided deep insights into the secular interactions of the planets. These are examined in more detail in Sect. 9.10.