Math. Proc. Camb. Phil. Soc. (1975), **77**, 159
With 1 *text-figure*
MPCPS 77–13
Printed in Great Britain

Almost-conservative second-order differential equations

By H. P. F. SWINNERTON-DYER
St Catharine's College, Cambridge

(*Received* 3 *July* 1974)

During the last thirty years an immense amount of research has been done on differential equations of the form

$$\ddot{x} + \omega^2 x = \epsilon f(x, \dot{x}, t) \tag{1}$$

where $\epsilon > 0$ is small. It is usually assumed that the perturbing term on the right-hand side is a 'good' function of its arguments and that its dependence on t is purely trigonometric; this means that there is an expansion of the form

$$f(x, \dot{x}, t) = a_0(x, \dot{x}) + \sum_1^\infty \{a_n(x, \dot{x}) \cos \omega_n t + b_n(x, \dot{x}) \sin \omega_n t\} \tag{2}$$

where the ω_n are constants, and that one may impose any conditions on the rate of convergence of the series which turn out to be convenient. Without loss of generality we can assume

$$\omega_n > 0 \quad \text{for all} \quad n > 0,$$

and for convenience we shall sometimes write $w_0 = 0$. Often f is assumed to be periodic in t, in which case it is implicit that the period is independent of x and $\dot{x}$. (We can also allow f to depend on ϵ, provided it does so in a sensible manner.)

Under reasonable hypotheses, a trajectory of (1) remains during any time $O(1)$ within a distance $O(\epsilon)$ of the corresponding trajectory of the equation

$$\ddot{x} + \omega^2 x = 0, \tag{3}$$

which is obtained by neglecting the perturbing term in (1). The trajectories of (3) have the form

$$\dot{x}^2 + \omega^2 x^2 = \text{constant}$$

and form a family of ellipses in the $x, \dot{x}$ plane; so we should think of any trajectory of (1) as drifting slowly across this family of ellipses. The principal problem of the theory is to determine, for a given equation and initial conditions, on which ellipse or band of ellipses a trajectory eventually settles down – the answer to this being almost independent of ϵ provided ϵ is small enough. Both the methods used and the results obtained depend crucially on the fact that every trajectory of (3), whatever its amplitude, has the same frequency ω. One well-known consequence of this is that the only terms on the right-hand side of (2) which can materially influence the eventual motion are the t-independent term and the terms for which ω_n is approximately an integral multiple of the free frequency ω. In other words, with eventual error $o(1)$ one can ignore the non-resonant terms in (2).

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

It is natural to ask how far similar considerations apply to equations of the form

$$\ddot{x} + g(x) = \epsilon f(x, \dot{x}, t), \tag{4}$$

where $g(x)$ is no longer linear but still satisfies

$$g(x) \operatorname{sign} x > 0 \quad \text{for} \quad x \neq 0. \tag{5}$$

For convenience we write

$$G(x) = \int_0^x g(\xi)\, d\xi$$

so that $G(x) > 0$ for $x \neq 0$, by (5); and we make the additional hypothesis

$$G(x) \to \infty \quad \text{as} \quad |x| \to \infty. \tag{6}$$

Just as with (1), under reasonable hypotheses a trajectory of (4) remains during any time $O(1)$ within a distance $O(\epsilon)$ of the corresponding trajectory of the equation

$$\ddot{x} + g(x) = 0. \tag{7}$$

The trajectories of (7) have the form

$$\dot{x}^2 + 2G(x) = \text{const.} \tag{8}$$

so that in view of (6) they form a family of ovals in the $x, \dot{x}$ plane. The crucial difference from (1) is that now the frequency of a trajectory of (7) does depend on its amplitude. Thus instead of a summand in (2) being always resonant or always non-resonant, which is what happens for (1), it should now be thought of as resonant at a certain discrete set of amplitudes and non-resonant elsewhere.

In the almost-linear case (1) it is well known that resonant terms in f are capable of counterbalancing the drift across the family of trajectories caused by the term $a_0(x, \dot{x})$. In particular, if the 'averaged' equation

$$\ddot{x} + \omega^2 x = \epsilon a_0(x, \dot{x}) \tag{9}$$

admits sustained oscillations and if f is periodic in t with frequency ω, then in most physically interesting cases the eventual behaviour of the solution of (1) differs substantially from that of the solutions of (9). Quite often the behaviour is even qualitatively different. It is tempting to hope that some such effect occurs with (4)– specifically, to hope that any sufficiently important harmonic in (2) is capable of imposing its own period on the solutions of (4) and thereby also determining their eventual amplitude. In the case when f is periodic, which is probably reasonably typical, the first step is to find out under what conditions (4) admits a periodic solution with the same period as f. General results on this problem have been obtained by Loud(3), and very detailed investigations of a number of particular cases have been made by Obi; their work shows that such periodic solutions will very often exist and may even be asymptotically stable. So the situation is this; the solutions of (4) will drift across the family of orbits of (7) in much the same way as the solutions of the 'averaged' equation

$$\ddot{x} + g(x) = \epsilon a_0(x, \dot{x}) \tag{10}$$

except in narrow bands about certain critical amplitudes, at which there is resonance between the corresponding solution of (7) and a sufficiently important summand in (2).

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

When f is periodic, the methods of Loud show that any such band contains periodic solutions of (4); and for general f it is proved in this paper that (under some additional conditions) to any such band there correspond solutions of (4) which remain in that band throughout $-\infty < t < \infty$. But this does not answer the main question: does the general solution of (4) get trapped in one of these bands, or does it drift across every such band and therefore have the same eventual behaviour as the general solution of (10)?

The main purpose of this paper is to show that it is the second alternative which happens, so that the trajectories which remain permanently in one of the exceptional bands represent an untypical rather than a typical eventual motion. Indeed, the proportion of trajectories which remain in an exceptional band tends to 0 with ϵ, though it may be non-zero for a given $\epsilon > 0$. For simplicity, this will be proved only under additional hypotheses which are, roughly speaking, that two unlikely events do not happen simultaneously. The result is primarily of interest for those equations (4) for which every solution is eventually uniformly bounded– that is, for which there is a constant B depending on f and g but not on ϵ such that each solution of (4) satisfies

$$|x(t)| \leqslant B, \quad |\dot{x}(t)| \leqslant B \quad \text{for all large enough positive } t.$$

Unfortunately one can at present draw such conclusions only under very restrictive hypotheses; for a result which contains most of the known useful ones, see (2), theorem 9. To get around this difficulty I therefore confine myself in this paper to parts of trajectories which lie in a region of the form $|x| \leqslant B$, $|\dot{x}| \leqslant B$ for some fixed B independent of ϵ.

In respect of ϵ, we adopt the convention which is standard in this subject. That is, any statement is subject to a condition of the form $0 < \epsilon < \epsilon_0$, where ϵ_0 is a constant which may depend on f and g and which can be rechosen smaller at any stage of the argument. Thus any qualitative statement is exactly true provided ϵ is small enough, rather than becoming more and more nearly true as ϵ tends to zero.

We assume that the family of closed curves (8) is parametrized by a sensibly chosen parameter a, which might for example be the total amplitude of the curve, or the constant on the right-hand side of (8). For convenience of description we take $a = 0$ to correspond to the singular point $x = \dot{x} = 0$ and positive values of a to correspond to the real closed curves (8); and we suppose that the constant on the right-hand side of (8) is a monotone increasing function of a. To any $a > 0$ there therefore corresponds a periodic solution of (7), whose frequency we denote by $\omega(a)$; and the general solution of (7) takes the form

$$x = X(a, (t+c)\,\omega(a))$$

for some constants a and c, where X is a well-determined function which is periodic in its second argument with period 2π. It is known that the differentiability properties of X and ω in $a > 0$ are at least as good as those of g. We denote partial derivatives of X with respect to its first and second arguments by subscripts $_1$ and $_2$; and we denote derivatives of ω, and of the function $A(a)$ defined in (12) below, by dashes. With this notation (7) becomes

$$\omega^2 X_{22} + g(X) = 0 \tag{11}$$

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

and (8) gives

$$\omega^2 X_2^2 + 2G(X) = 2A(a) \tag{12}$$

say, for since the left-hand side does not depend on t it cannot depend on c either. By hypothesis, $A(a)$ is non-negative and strictly increasing in $a \geqslant 0$; and it has a continuous first derivative in $a > 0$ because the left-hand side of (12) has. Indeed, differentiating (12) with respect to a we obtain

$$\omega\omega' X_2^2 + \omega^2 X_2\{X_{12} + (t+c)\,\omega' X_{22}\} + \{X_1 + (t+c)\,\omega' X_2\}\,g(X) = A', \tag{13}$$

which will be needed below.

Having set up this machinery, we can now turn to equation (4). By analogy with the standard method of treating equation (1), it is tempting to define new variables r, θ by means of the implicit equations

$$x = X(r, (t+\theta)\,\omega(r)), \quad \dot{x} = \omega(r)\,X_2(r, (t+\theta)\,\omega(r)),$$

and to replace (4) by a pair of first-order differential equations for r and θ. But this turns out to be unsatisfactory because, although the expression for $\dot{\theta}$ is small in the sense that ϵ is formally a factor of it, the presence of secular terms prevents it being treated as small in the usual way when $\omega' \neq 0$. Instead we choose new variables r, u defined by the implicit equations

$$x = X(r, u), \quad \dot{x} = \omega(r)\,X_2(r, u). \tag{14}$$

To show that this and continuity are enough to determine r and u uniquely for given x and $\dot{x}$, it is enough to show that the Jacobian of the inverse transformation (14) is non-zero. But this Jacobian is

$$\begin{aligned}\frac{\partial(x, \dot{x})}{\partial(r, u)} &= \begin{vmatrix} X_1 & \omega' X_2 + \omega X_{12} \\ X_2 & \omega X_{22} \end{vmatrix} \\ &= \omega X_1 X_{22} - \omega' X_2^2 - \omega X_2 X_{12} \\ &= \omega^{-2}\{\omega X_1 + u\omega' X_2\}\{\omega^2 X_{22} + g(X)\} - \omega^{-1} A'(r) = -\omega^{-1} A'(r)\end{aligned}$$

where we have used first (13) in the form

$$A'(r) = \omega\omega' X_2^2 + X_2\{\omega^2 X_{12} + \omega\omega' u X_{22}\} + \{X_1 + \omega^{-1}\omega' u X_2\}\,g(X)$$

and then (11). We therefore impose the further hypothesis $A'(a) > 0$ for $a > 0$ on the parametrization of the curves (8); this is not a restriction on the equation (4) since, as we have already remarked, one possible choice for a is A itself.

Now differentiating the first equation (14) with respect to t and comparing it with the second we obtain

$$\dot{r} X_1 + \dot{u} X_2 = \omega X_2;$$

and differentiating the second equation (14), substituting it into (4) and using (11) we obtain

$$\dot{r}(\omega' X_2 + \omega X_{12}) + \dot{u}\omega X_{22} = \omega^2 X_{22} + \epsilon f.$$

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

By a calculation like that of the previous paragraph, these two equations give

$$A'\dot{r} = \epsilon\omega X_2 f, \tag{15}$$

$$A'(\dot{u}-\omega) = -\epsilon\omega X_1 f. \tag{16}$$

In this paper we shall be concerned with the behaviour of trajectories within rings of the form $r_1 \leqslant r \leqslant r_2$, where r_1, r_2 are independent of ϵ and $r_1 > 0$. In any such ring we have

$$\dot{r} = O(\epsilon), \quad \dot{u} = \omega(r) + O(\epsilon)$$

by (15) and (16).

With the substitution (14) we can regard $X_2 f$ as a function of r, u and t; and as a function of u it is periodic with period 2π. Moreover its dependence on t is made explicit by (2). We therefore write

$$a_n(x, \dot{x})\, X_2(r, u) = a_{0n}(r) + \sum_{m=1}^{\infty} a_{mn}(r) \cos(mu + \theta_{mn}),$$

$$b_n(x, \dot{x})\, X_2(r, u) = b_{0n}(r) + \sum_{m=1}^{\infty} b_{mn}(r) \cos(mu + \phi_{mn});$$

and when convenient we shall use the convention $\theta_{0n} = \phi_{0n} = 0$ so as to give the first term on the right the same shape as the others. Thus formally (15) becomes

$$A'\dot{r} = \epsilon\omega\Sigma\Sigma\{a_{mn}\cos(mu+\theta_{mn})\cos\omega_n t + b_{mn}\cos(mu+\phi_{mn})\sin\omega_n t\}. \tag{17}$$

I shall assume that the double series

$$\Sigma\Sigma\{|a_{mn}(r)| + |b_{mn}(r)|\}$$

is convergent uniformly in any ring of the form $0 < r_1 \leqslant r \leqslant r_2$; for this it is certainly sufficient that the series (2) and its first and second derivatives with respect to x and $\dot{x}$ are all absolutely convergent uniformly in any bounded domain for x and $\dot{x}$. So this is just a well-behavedness condition on f, of a type which will be satisfied for any physically interesting equation. The 'averaged' form of (17) is

$$A'(r)\dot{r} = \epsilon\omega(r)\, a_{00}(r) \tag{18}$$

and by hypothesis $A' > 0$ and $\omega > 0$. So a solution of (18) increases when $a_{00}(r) > 0$ and decreases when $a_{00}(r) < 0$; and any solution except an unstable constant one either tends to zero or infinity or to a point at which $a_{00}(r)$ changes from positive to negative as r increases. (This assumes that $a_{00}(r)$ has only isolated zeros; the question what happens in an interval throughout which a_{00} vanishes is beyond the scope of this paper.) To show that the behaviour of solutions of (4) is substantially the same as that of (18) it will be enough to consider what happens in a ring in which $a_{00}(r)$ is non-zero and of constant sign – which without loss of generality we can take to be positive. So henceforth we assume that there exists $c > 0$ such that

$$a_{00}(r) \geqslant c > 0 \text{ for } r_1 \leqslant r \leqslant r_2; \tag{19}$$

for the rest of the paper we shall regard r_1 and r_2 as fixed. By hypothesis there is a finite

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

set S of pairs (m, n) such that

$$\Sigma\Sigma\{|a_{mn}(r)| + |b_{mn}(r)|\} < \tfrac{1}{2}c \quad \text{for} \quad r_1 \leqslant r \leqslant r_2 \tag{20}$$

where the sum is extended over all pairs (m, n) not in S other than $(0, 0)$; we make the convention that $(0, 0)$ is not in S.

Now let $M > 0$ be a large fixed number, depending on f and g but not on ϵ; the conditions which M has to satisfy will be listed below as they become relevant to the argument. We note that ϵ_0, the upper bound of values of ϵ for which the argument is valid, will depend on the choice of M. To each pair (m, n) in S we associate the 'resonance zone'

$$|m\omega(r) - \omega_n| \leqslant M\epsilon^{\frac{1}{2}}$$

of values of r. For the purpose of our description we shall suppose that to each pair (m, n) in S corresponds only a single resonance zone and this zone has width $o(1)$; this is true for example whenever $\omega(r)$ is strictly monotone in $r_1 \leqslant r \leqslant r_2$. By choosing ϵ_0 small enough we can ensure that every resonance zone within $r_1 \leqslant r \leqslant r_2$ has $m \neq 0$ and $n \neq 0$, and that two resonance zones with different values of ω_n/m do not overlap. There can of course be several pairs (m, n) in S with the same value of ω_n/m, but in such a case we need only worry about the resonance zone corresponding to the smallest value of m.

We next consider the behaviour of a trajectory in a time interval $\tau \leqslant t \leqslant \tau + T$ under the hypothesis that during this interval it remains entirely inside $r_1 \leqslant r \leqslant r_2$ and entirely outside every resonance zone. By (19) and (20) we have

$$r(\tau + T) - r(\tau) = \int_\tau^{\tau+T} \dot{r}\, dt \geqslant \tfrac{1}{2}\epsilon cT \min(\omega/A')$$

$$+ \epsilon \sum_S \sum \int_\tau^{T+\tau} \frac{\omega}{A'} \{a_{mn} \cos(mu + \theta_{mn}) \cos \omega_n t + b_{mn} \cos(mu + \phi_{mn}) \sin \omega_n t\}\, dt$$

and we need to find a bound for these integrals. Each such integral is a sum of four terms, of which a typical one is

$$\int_\tau^{T+\tau} \frac{\omega}{2A'} a_{mn} \cos(mu - \omega_n t + \theta_{mn})\, dt$$

$$= \int_\tau^{\tau+T} \frac{\omega a_{mn}}{2(mA'\omega - A'\omega_n - m\epsilon\omega X_1 f)} \left(m\omega - \omega_n - \frac{m\epsilon\omega X_1 f}{A'}\right) \cos(mu - \omega_n t + \theta_{mn})\, dt$$

$$= \left[\frac{\omega a_{mn}}{2(mA'\omega - A'\omega_n - m\epsilon\omega X_1 f)} \sin(mu - \omega_n t + \theta_{mn})\right]_\tau^{\tau+T}$$

$$+ \int_\tau^{\tau+T} \epsilon O((mA'\omega - A'\omega_n - m\epsilon\omega X_1 f)^{-2})\, dt = O(M^{-1}\epsilon^{-\frac{1}{2}} + M^{-2}T)$$

since we are outside any resonance zone and therefore

$$|m\omega - \omega_n - m\epsilon\omega X_1 f/A'| \geqslant \tfrac{1}{2}M\epsilon^{\frac{1}{2}}$$

provided ϵ is small enough. Here the constant implicit in the 'O' notation depends on

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

f, g, r_1 and r_2 but not on the choice of M or ϵ. Substituting back into the previous equation we obtain

$$\begin{aligned} r(\tau+T)-r(\tau) &\geqslant \tfrac{1}{2}\epsilon cT \min(\omega/A') + O(M^{-1}\epsilon^{\frac{1}{2}} + M^{-2}\epsilon T) \\ &\geqslant \tfrac{1}{4}\epsilon cT \min(\omega/A') \quad \text{for} \quad T \geqslant \epsilon^{-\frac{1}{2}} \end{aligned}$$

provided we choose M large enough. Moreover (15) implies

$$r(\tau+T)-r(\tau) = O(\epsilon^{\frac{1}{2}}) \quad \text{if} \quad 0 \leqslant T < \epsilon^{-\frac{1}{2}}. \tag{21}$$

Thus if a trajectory is initially in $r_1 + O(\epsilon^{\frac{1}{2}}) \leqslant r \leqslant r_2$ and outside any resonance zone, then r increases (subject to minor fluctuations) until either $r = r_2$ or the trajectory enters a resonance zone.

It only remains to study the behaviour of trajectories in or near a particular resonance zone, corresponding say to the pair (m_0, n_0) in S. We define r_0 by

$$m_0\omega(r_0) - \omega_{n_0} = 0,$$

so that $r = r_0$ can be regarded as the centre of the resonance zone; and we make the further hypothesis

$$\omega'(r_0) \neq 0. \tag{22}$$

We are now concerned with trajectories which start in the resonance zone, and we study them only so long as they remain in the 'extended resonance zone'

$$|m_0\omega(r) - \omega_{n_0}| \leqslant 2M\epsilon^{\frac{1}{2}}. \tag{23}$$

Provided M is large enough, the argument of the previous paragraph shows that such a trajectory can only leave the extended resonance zone in the direction of increasing r, and that it will take at least a time comparable with $M\epsilon^{-\frac{1}{2}}$ to do so. This is the last condition which we need to impose on M, so henceforth we can regard M as fixed; this means that the constants implicit in the 'O' notation may depend on the choice of M. It follows from (22) and (23) that within the extended resonance zone

$$r = r_0 + O(\epsilon^{\frac{1}{2}}),$$

so we replace r by the new variable ρ defined by

$$r = r_0 + \epsilon^{\frac{1}{2}}\rho;$$

this means that $\rho = O(1)$ for the stretches of trajectory which we are now concerned with. By (15), the differential equation for ρ is

$$A'(r)\dot{\rho} = \epsilon^{\frac{1}{2}}\omega(r)\, X_2 f. \tag{24}$$

In the arguments of the trigonometric functions it is convenient to replace u by

$$\sigma = m_0 u - \omega_{n_0} t;$$

by (16) this satisfies the differential equation

$$\begin{aligned} A'(r)\,\dot{\sigma} &= A'(r)\{m_0\omega(r) - \omega_{n_0}\} - m_0\epsilon\omega(r)\, X_1 f \\ &= \epsilon^{\frac{1}{2}}A'(r)\,\omega'(r)\,\rho + O(\epsilon). \end{aligned} \tag{25}$$

Now in (24) we can regard X_2 and f as functions of σ, t and r; and we can replace

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

r by r_0 in (24) and (25) at the expense of a term $O(\epsilon)$ on the right of (24). Over a time $O(\epsilon^{-\frac{1}{2}})$ the terms $O(\epsilon)$ on the right of (24) and (25) will only make a difference of $O(\epsilon^{\frac{1}{2}})$ to the behaviour of trajectories in the ρ, σ space; so we can drop these terms. With these modifications our system of equations becomes

$$\dot{\rho} = \epsilon^{\frac{1}{2}}\{\omega(r_0)/A'(r_0)\}\,\Sigma\,\Sigma\, c_{mn}\cos(mm_0^{-1}\sigma + \lambda_{mn}t + \psi_{mn}), \tag{26}$$

$$\dot{\sigma} = \epsilon^{\frac{1}{2}}\omega'(r_0)\,\rho, \tag{27}$$

where we make the convention that $\lambda_{mn} = 0$ if and only if $n = 0$. Note here that

$$c_{00} = a_{00}(r_0) > 0, \psi_{00} = 0. \tag{28}$$

The equations (26) and (27) are in the standard form of orthodox small parameter theory, with $\epsilon^{\frac{1}{2}}$ in place of the usual ϵ. Over a time $O(\epsilon^{-\frac{1}{2}})$ we can therefore replace them by their time-averaged form with error $o(1)$; for a rigorous proof of this see (1), chapter 6. Thus over a time $O(\epsilon^{-\frac{1}{2}})$ we may with error $o(1)$ replace the equations (24) and (25) by

$$\dot{\rho} = \epsilon^{\frac{1}{2}}\{\omega(r_0)/A'(r_0)\}\,\Sigma c_{m0}\cos(mm_0^{-1}\sigma + \psi_{m0}), \tag{29}$$

$$\dot{\sigma} = \epsilon^{\frac{1}{2}}m_0\omega'(r_0)\,\rho; \tag{30}$$

and we can make the constant implicit in the $O(\epsilon^{-\frac{1}{2}})$ as large as we like by choosing ϵ_0 small enough.

These equations are easy to deal with. Write

$$h(\sigma) = \sum_0^\infty c_{m0}\cos(mm_0^{-1}\sigma + \psi_{m0}),$$

$$H(\sigma) = \sum_1^\infty m_0 m^{-1} c_{m0}\sin(mm_0^{-1}\sigma + \psi_{m0}),$$

$$B = 2\omega(r_0)/m_0 A'(r_0)\,\omega'(r_0);$$

then we have

$$2\rho\,d\rho = Bh(\sigma)\,d\sigma,$$

which can be integrated at once to give the equation of the trajectories in the form

$$\rho^2 = Bc_{00}\sigma + BH(\sigma) + \text{const.} \tag{31}$$

For convenience of description we assume $\omega'(r_0) > 0$ and so $B > 0$, and we regard σ as measured horizontally and ρ vertically. The following arguments will depend crucially on the fact that any curve (31) is symmetric about the horizontal axis $\rho = 0$. For convenience of description we assume $\omega'(r_0) > 0$ and so vertically. The following arguments will depend crucially on the fact that any curve (31) is symmetric about the horizontal axis $\rho = 0$.

Any trajectory of (29) and (30) moves to the left as long as $\rho < 0$, and to the right while $\rho > 0$. The singular points of the system are given by

$$\rho = h(\sigma) = 0$$

and we shall at least assume that there are only finitely many of them up to periodicity; this is reasonable since we know that the constant term in $h(\sigma)$ is non-zero. Consider a trajectory which starts in $\rho < 0$. So long as it remains in that region σ decreases and

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

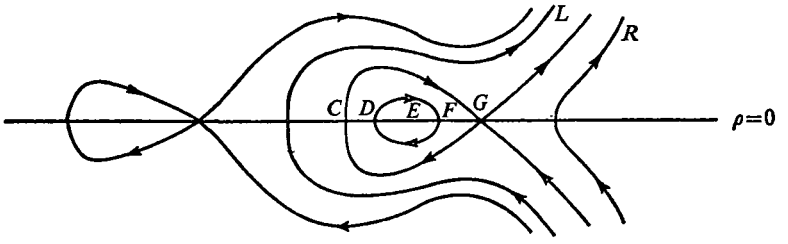
hence ρ remains bounded, by (31) and the fact that $H(\sigma)$, being periodic and continuous, is bounded; moreover σ is bounded below since the right-hand side of (31) cannot go negative. Thus the trajectory either eventually crosses $\rho = 0$ or else tends to a singular point which must be on $\rho = 0$; and the second alternative can only happen if the constant in (31) has one of the discrete set of values which correspond to singular points. If a trajectory reaches $\rho = 0$ then by symmetry it continues along the mirror image of its previous path. In particular, with the exception of a discrete set of trajectories, a trajectory which starts on $\rho = -2M/m_0\omega'(r_0)$, which corresponds to the lower edge of the extended resonance zone, will eventually reach $\rho = 2M/m_0\omega'(r_0)$ which corresponds to the upper edge of the extended resonance zone. Moreover it will normally not pass near a singular point, and will therefore cross this strip in a time $O(\epsilon^{-\frac{1}{2}})$. Indeed if we make the additional hypothesis that the zeros of $h(\sigma)$ are all simple, then a trajectory can only fail to cross this zone in a time $O(\epsilon^{-\frac{1}{2}})$ if it passes near a singular point, and this will only happen if the constant in (31) is close to a constant corresponding to a singular point. Hence with the exception of a subset of small measure, a trajectory of (29) and (30) will cross the extended resonance zone in a time $O(\epsilon^{-\frac{1}{2}})$; and so by our previous results the same is true for the trajectories of the true equations (24) and (25). Thus all but a proportion $o(1)$ of trajectories cross the resonance zone successfully, as was asserted at the beginning of this paper.

However, it is not hard to show that some trajectories must be trapped in the resonance zone provided $h(\sigma)$ does take negative values. For convenience of description we assume that the zeros of $h(\sigma)$ are all simple, though the result does not depend on this hypothesis. A singular point is a saddle point if $h(\sigma)$ is monotone increasing there, so that the right-hand side of (31) has a local minimum. It is a focus, and by symmetry actually a centre, if $h(\sigma)$ is monotone decreasing there so that the right-hand side of (31) has a local maximum. Let G be a saddle point; then the separatrix which leaves G moving downwards and to the left must either tend to another singular point or cross $\rho = 0$, by an argument much like that of the previous paragraph. Moreover if G is such that the values of $\{c_{00}\sigma + H(\sigma)\}$ to the right of G are all strictly greater than its value at G – and it follows from (28) that such points exist – the separatrix which leaves G upwards and to the right cannot tend to any point of $\rho = 0$ nor cross $\rho = 0$, by (31); hence it stays in $\rho > 0$ and has $\sigma \to +\infty$ by the same type of argument. The loop of the separatrix to the left of G (which is a loop because of symmetry) must contain at least one centre, which we call E. The simplest possible configuration, and a sufficiently typical one to illustrate the argument, is as shown in the figure below. Here C is the point at which the loop of the separatrix crosses $\rho = 0$ again, E is the centre, and D and F are the left- and right-hand points of some trajectory which lies entirely inside the loop of the separatrix.

The figure shows trajectories of (29) and (30); but trajectories of the true equations (24) and (25) behave in much the same way except that of course they need not form closed curves. In particular the following two properties are now obvious:

(i) If a trajectory of the true equations passes through a point of the closed interval DF then it either stays for ever in the resonance zone (and indeed inside the separatrix loop) or else it subsequently crosses $\rho = 0$ to the left of D.

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press



(ii) If a trajectory of the true equations which starts below $\rho = 0$ and eventually leaves the resonance zone at its upper edge meets the line $\rho = 0$ to the left of D, then it has at least one strict crossing of $\rho = 0$ to the left of D.

Now fix a time t_0 and let P denote any point of the resonance zone below $\rho = 0$; and consider the trajectory of the true equations passing through P at time t_0. We say that P belongs to the set $\mathscr{L}$ (respectively $\mathscr{R}$) if the trajectory eventually leaves the extended resonance zone at its upper edge and if it does (respectively does not) meet the line $\rho = 0$ at least once strictly to the left of D. If P belongs to $\mathscr{L}$ then its associated trajectory actually crosses $\rho = 0$ strictly to the left of D; so by a standard continuity argument $\mathscr{L}$ is open. If P belongs to $\mathscr{R}$ then the associated trajectory cannot pass through D; so, by a continuity argument again, $\mathscr{R}$ is open. The trajectories labelled L and R in the figure, which are exact for (29) and (30) and thus approximate for the true equations, are enough to show that neither $\mathscr{L}$ nor $\mathscr{R}$ is empty. Since $\mathscr{R}$ and $\mathscr{L}$ are disjoint, their union cannot be the whole resonance zone below $\rho = 0$, that being a connected set. Hence there is a trajectory which is in the resonance zone at time t_0 and remains in the extended resonance zone for ever after. Now repeat this argument with $t_0 - N$ instead of t_0 (for $N = 0, 1, 2, \ldots$) and follow the resulting trajectory for a time N; thus for each N we obtain a point P_N in the extended resonance zone such that the trajectory which passes through P_N at time t_0 remains in the extended resonance zone for all times from $t_0 - N$ onwards. Since the equations (24) and (25) are periodic with period $2\pi m_0$ in σ, we can regard the extended resonance zone as compact; so it contains a point P^* which is a point of accumulation of the P_N. Clearly the trajectory which passes through P^* at time t_0 lies in the extended resonance zone for all time, and is thus an exception to the general upwards drift. (We can also find a trajectory which enters the resonance zone from below at time t_0, and stays in the extended resonance zone thereafter, by considering the intersections of $\mathscr{R}$ and $\mathscr{L}$ with the lower edge of the resonance zone and using the same argument as before.)

By a more sophisticated argument of the same kind, we can even show that there is a trajectory which remains close to G for all time. For let C_1, C_2 denote circles with centre G and radii $\epsilon^{\frac{1}{3}}$ and $2\epsilon^{\frac{1}{3}}$ respectively. By an argument much like that of the previous paragraph, if a trajectory is inside C_1 at time t_0 and subsequently crosses C_2, its first strict crossing of C_2 must be within one of two disjoint subarcs of C_2, one on its lower left-hand part and the other on its upper right-hand part; and each possibility can occur, by starting at a point near the relevant outward separatrix and not too close to G. These two possibilities correspond to two disjoint non-empty open subsets of the interior of C_1, and since that interior is connected it must contain points which belong to neither of the subsets. A trajectory which passes through such a point at

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press

time t_0 will never thereafter pass outside C_2. By a point-of-accumulation argument like that of the previous paragraph, it follows that there is a trajectory which remains inside C_2 for all time. I have not been able to apply any argument of this kind to prove that there is a trajectory which always remains near the centre-like point E.

Stronger results can be obtained by making use of the variational equations. In particular I can show that there is exactly one trajectory which remains always near G, though there may be other trajectories which always remain in the resonance zone. If moreover the original equation (4) is periodic then there is just one trajectory which is periodic of the same period and always remains near E. But the methods used for this proof are more naturally applicable to classical small parameter theory, and are therefore postponed to a later paper.

REFERENCES

(1) Bogoliubov, N. N. and Mitropolsky, Y. A. *Asymptotic methods in the theory of non-linear oscillations* (Moscow, 1955; English translation, Delhi, 1961).

(2) Cartwright, M. L. and Swinnerton-Dyer, H. P. F. Boundedness theorems for some second order differential equations, I. *Ann. Math. Pol.* **29** (1974), 229–254.

(3) Loud, W. S. Periodic solutions of perturbed second-order autonomous equations. *Mem. Amer. Math. Soc.* **47** (1964).

https://doi.org/10.1017/S0305004100049495 Published online by Cambridge University Press