Ergod. Th. & Dynam. Sys. (1988), **8***, 119–138
Printed in Great Britain

Instability of standing waves for non-linear Schrödinger-type equations

CHRISTOPHER K. R. T. JONES
Department of Mathematics, University of Maryland, College Park, Maryland 20742, *USA*

Abstract. A theorem is proved giving a condition under which certain standing wave solutions of non-linear Schrödinger-type equations are linearly unstable. The eigenvalue equations for the linearized operator at the standing wave can be analysed by dynamical systems methods. A positive eigenvalue is then shown to exist by means of a shooting argument in the space of Lagrangian planes. The theorem is applied to a situation arising in optical waveguides.

1. *Introduction*

Non-linear Schrödinger equations arise in optics as slowly varying envelope approximations for the non-linear wave equation (see [**13**] and [**16**]). In this paper we shall focus on the case where the non-linear term is spatially dependent and the space variable is one-dimensional:

$$i\frac{\partial\phi}{\partial t}=\frac{\partial^2\phi}{\partial x^2}+[\beta+f(x,|\phi|^2)]\phi, \tag{1.1}$$

where $i=\sqrt{-1}$, β is a real number and f is at least C^3. The number β is called the propagation constant. These equations are usually written without the β present. Notice that if $u(x, t)$ satisfies (1.1), then $e^{-i\beta t}u(x, t)$ will satisfy the equation without the β term. A standing wave u is a real time-independent solution of (1.1) for some β; it therefore satisfies

$$u_{xx}+[\beta+f(x,|u|^2)]u=0. \tag{1.2}$$

We shall address the question of the stability of a solution of (1.2) relative to the evolution equation (1.1). We shall describe a mechanism for the instability of certain standing waves. The instability will be determined by linearizing (1.1) at the standing wave in question.

The main theorem of this paper holds under fairly general hypotheses on the non-linearity f.

(H1) $f:\mathbb{R}^2\to\mathbb{R}$ is at least C^3 and all derivatives up to order 3 are bounded on a set of the form $\mathbb{R}\times U$, where U is a neighbourhood of $0\in R$.

(H2) $f(x, 0)\to 0$ exponentially as $x\to\pm\infty$.

Write (1.1) in real and imaginary parts, $\phi=v+iw$:

$$\begin{aligned} v_t&=w_{xx}+f(x, v^2+w^2)w+\beta w,\\ w_t&=-v_{xx}-f(x, v^2+w^2)v-\beta v. \end{aligned} \tag{1.3}$$

Suppose $u(x)$ is a standing wave solution; then it is a real ($w=0$) time-independent solution of (1.3). We shall assume the following about $u(x)$.

(H3) $u(x)$ is C^2 and decays exponentially as $x\to\pm\infty$.

The existence of such a $u(x)$ does not necessarily follow from the hypotheses (H1) and (H2). The main application will be to optical waveguides (see § 5) and the theorem will be applied to standing waves that are found by phase portrait techniques. A condition that is necessary for $u(x)$ to decay exponentially is that $\beta<0$, so we shall make this a hypothesis.

(H4) $\beta<0$.

The linearization of (1.3) at $u(x)$ is formally

$$\begin{aligned} p_t &= q_{xx}+f(x,u^2)q+\beta q,\\ q_t &= -p_{xx}-f(x,u^2)-2D_2f(x,u^2)u^2p-\beta p, \end{aligned} \tag{1.4}$$

where D_2f is the derivative with respect to its second variable. Setting

$$\begin{aligned} L_- &= \frac{d^2}{dx^2}+f(x,u^2)+\beta,\\ L_+ &= \frac{d^2}{dx^2}+f(x,u^2)+2D_2f(x,u^2)u^2+\beta, \end{aligned} \tag{1.5}$$

(1.4) can be rewritten as the system (this is the notation of [17])

$$\begin{pmatrix} p\\ q\end{pmatrix}_t=\begin{pmatrix} 0 & L_-\\ -L_+ & 0\end{pmatrix}\begin{pmatrix} p\\ q\end{pmatrix}=N\begin{pmatrix} p\\ q\end{pmatrix}. \tag{1.6}$$

We shall study instability by finding unstable eigenvalues of the operator N, which is natural to consider as an operator on $H^1(\mathbb{R})\times H^1(\mathbb{R})$.

The operators L_+ and L_- are both self-adjoint on $H^1(\mathbb{R})$ and therefore their spectra are real. The spectrum of N will definitely not be real; in fact much of it will lie on the imaginary axis. The interesting aspect of this problem is the relationship between the spectrum of N and the spectra of L_+ and L_-. Using the hypotheses made, it can be shown (see § 4) that the spectra of both L_+ and L_- in $\{\lambda>0\}$ consist of finitely many eigenvalues. The following quantities are therefore well defined:

$$P=\text{number of positive eigenvalues of } L_+$$

$$Q=\text{number of positive eigenvalues of } L_-.$$

The main theorem we shall prove in this paper relates P and Q to the existence of positive eigenvalues of N.

THEOREM 1. *If $P-Q\neq 0, 1$, there is a real positive eigenvalue of the operator N.*

From Sturm–Liouville theory, P and Q can be determined by considering solutions of $L_+v=0$ and $L_-v=0$ respectively. In fact P and Q equal the number of zeros of the associated solution v. Notice that $L_-v=0$ is actually satisfied by the standing wave itself. It follows that

$$Q=\text{number of zeros of the standing wave } u(x).$$

For positive solutions, theorem 1 then says that if $P>1$, the standing wave is unstable. It is known that if f is independent of x and of the appropriate form, for instance $f(|u|) = |u|^\sigma$, the ground state is stable (see [7] and [17]). In this case the ground state is the unique positive solution and it is easy to calculate that $P = 1$, hence $P - Q = 1$. The spatial dependence of f, however, offers the possibility of unstable waves.

The operator L_+ is the derivative of the non-linear operator which is the right-hand side of (1.1). A change in P, as some parameter varies, would therefore indicate that bifurcation of standing waves had taken place. Theorem 1 then shows which of the bifurcating solutions are necessarily unstable. There are recent instability results of Berestycki and Lions [**4**] and Shatah and Strauss [**14**], but these are a different phenomenon.

A case of space-dependent f arises in optical waveguides, (see [**1**], [**15**] and [**11**]). In the last section of this paper we shall apply theorem 1 to conclude a new result on the instability of certain waves, as announced in [**11**].

The basic idea in the proof is to apply a dynamical systems point of view to the eigenvalue equations. The existence of eigenvalues is established by a shooting argument in the space of Lagrangian planes.

In the next section we shall set up the eigenvalue problem and show how it can be reformulated as the problem of finding a connecting orbit in a Grassmannian manifold. We call it a shooting argument for the following reasons. An orbit is determined for each real value of λ that satisfies the desired condition at $-\infty$. The boundary condition at $+\infty$ is formulated by requiring this orbit to 'connect' to a certain set. As λ varies, the topology of this orbit can be seen to change, and at this change the orbit is forced to 'connect' to the desired set, thus creating an eigenvalue.

The special structure of the equation leads to a restriction to the invariant submanifold of the Grassmannian consisting of the planes that are Lagrangian. The space of Lagrangian planes has fundamental group $\mathbb{Z}$. Some notion of winding therefore exists for trajectories in this submanifold; this is the Maslov index. This winding number supplies the topological measurement needed for the shooting argument described above. In this low-dimensional case a very explicit representation of the space of Lagrangian planes can be given (see the Appendix) which allows one to visualize the winding in a very concrete way. This aspect is important, since the Maslov index cannot be used directly because the trajectories of interest are not necessarily closed. The description is also rather satisfying and may be of independent interest.

This use of the Maslov index as a mechanism to determine eigenvalues has been known for some time as a higher-dimensional generalization of Sturmian theory. Indeed, according to one interpretation this idea lies at the heart of the Morse index theorem on geodesics. These cases are usually associated to periodic problems, but the application of the Maslov index to non-periodic problems has also appeared, (see [3], [**6**] and [**8**]). The situation presented in this paper, however, differs fundamentally from each of the above in that the Maslov index is not monotone in the (eigenvalue) parameter here. An exact count of the relevant eigenvalues is therefore not afforded by the technique, but a topological argument still gives the existence of these eigenvalues.

A framework for the consideration of the eigenvalue problem is set up in § 2. In § 3 the main lemma is proved that forms the foundation of the shooting argument. The proof is then completed in § 4 and the application to optical waveguides is worked out in § 5.

2. *Formulation of the eigenvalue problem*

Consider the eigenvalue equations of N (see (1.6)):

$$L_{-}q = \lambda p, \qquad L_{+}p = -\lambda q. \tag{2.1}$$

We shall restrict to the case $\lambda \in \mathbb{R}$ since the theorem asserts the existence of real eigenvalues. These are the ordinary differential equations

$$\begin{aligned} q'' + f(x, u^2)q + \beta q &= \lambda p, \\ p'' + f(x, u^2)p + 2D_2 f(x, u^2)u^2 p + \beta p &= -\lambda q. \end{aligned} \tag{2.2}$$

If we write (2.2) as a system, we obtain

$$p' = y, \qquad q' = w, \qquad y' = -g(x)p - \lambda q, \qquad w' = -h(x)q + \lambda p, \tag{2.3}$$

where

$$\begin{aligned} g(x) &= f(x, u^2) + \beta, \\ h(x) &= f(x, u^2) + 2D_2 f(x, u^2)u^2 + \beta, \end{aligned}$$

recalling that $u = u(x)$.

The system (2.3) is not Hamiltonian but becomes so if the change of variables $s = -q$ is made in (2.3). The equations are then

$$\left.\begin{aligned} p' &= y \\ s' &= -w \\ y' &= -g(x)p + \lambda s \\ w' &= h(x)s + \lambda p \end{aligned}\right\} = \left(\begin{array}{cc|cc} \multicolumn{2}{c|}{0} & 1 & 0 \\ & & 0 & -1 \\ \hline -g & \lambda & \multicolumn{2}{c}{0} \\ \lambda & h & & \end{array}\right)\begin{pmatrix} p \\ s \\ y \\ w \end{pmatrix}. \tag{2.4}$$

Theorem 1 will be proved by finding solutions of (2.4) that decay at $\pm\infty$ with $\lambda > 0$. This leads to the first formulation of the eigenvalue condition. Rewrite (2.4) with $P = (p, s, y, w)$ as

$$P' = A(\lambda, x)P. \tag{2.5}$$

LEMMA 2.1. *If* (2.5) *admits a solution* $P(x)$ *that decays exponentially as* $x \to \pm\infty$, *then* λ *is an eigenvalue of* N.

Proof. If $P(x)$ decays exponentially, then so does $p(x)$ and so do their derivatives. Since $q(x) = -s(x)$, $(p(x), q(x)) \in H^1(\mathbb{R}) \times H^1(\mathbb{R})$.

To make the boundary conditions as $x \to \pm\infty$ clearer, we shall compactify the equations (2.4) in x. In order to view the behaviour as $x \to \pm\infty$, set

$$x = \frac{1}{2\kappa} \ln\left(\frac{1+\tau}{1-\tau}\right). \tag{2.6}$$

This formula for compactification is due to Terman. The number κ will be chosen by lemma 2.2. Replacing x in its role as a dependent variable, but not as the independent variable, (2.5) becomes

$$\begin{aligned} p' &= y, \qquad s' = -w, \\ y' &= -g^c(\tau)p + \lambda s, \\ w' &= -h^c(\tau)s + \lambda p, \\ \tau' &= \kappa(1-\tau^2), \end{aligned} \tag{2.7}$$

where $' = d/dx$. Using the notation introduced above, (2.7) can be abbreviated as

$$P' = A^c(\lambda, \tau)P, \qquad \tau' = \kappa(1-\tau^2), \tag{2.8}$$

where

$$A^c(\lambda, \tau) = A\left(\lambda, \frac{1}{2\kappa}\ln\left(\frac{1+\tau}{1-\tau}\right)\right).$$

To recover a solution of (2.5), choose a solution of (2.8) that satisfies $P(x_0) = P_0$ and $\tau(x_0) = \tau_0$ with $\tau_0 \in (-1, +1)$. Set x_0 so that

$$\tau_0 = \frac{e^{2kx_0} - 1}{e^{2kx_0} + 1};$$

then $P(x)$ will be a solution of (2.5) satisfying $P(x_0) = P_0$.

The equation (2.8) is defined on $\mathbb{R}^4 \times (-1, +1)$. To extend to $\tau = \pm 1$, it must be checked that g^c and h^c can be smoothly extended. This follows from the hypotheses made on f, namely (H1) and (H2).

LEMMA 2.2. *The number κ can be chosen so that g^c and h^c can be extended to $\tau = \pm 1$ in a C^1 fashion.*

Proof. Firstly, consider $g(x) = f(x, u^2) + \beta$. Define $g^c(\pm 1) = \beta$ and without loss of generality assume that $\beta = 0$. To show that g^c is continuous, it suffices to note that

$$f(x, u^2) = f(x, u^2) - f(x, 0) + f(x, 0)$$

and

$$|f(x, u^2) - f(x, 0)| \le |D_2 f(x, \hat{u}^2(x))|\,|u(x)|^2 \tag{2.9}$$

for some $\hat{u}(x)$ with $|\hat{u}(x)| \le |u(x)|$, where $D_2 f$ denotes the partial derivative of f with respect to its second variable. (H1) and (H2) now imply that $f(x, u^2(x)) \to 0$ as $x \to \pm\infty$, in fact exponentially.

To see that g^c can be differentiated at $\tau = \pm 1$, it must be shown that

$$\frac{1}{1-\tau} g(x(\tau)) = \frac{1 + e^{2\kappa x}}{2} g(x)$$

tends to 0. But this follows from the above by choosing κ so that $e^{2\kappa x} f(x, u^2(x)) \to 0$. For g^c to be C^2, $\partial g/\partial x$ must decay exponentially; this follows from the bounds on the second derivatives. It can be shown similarly that h^c is extendable by using the bounds on some of the third partial derivatives.

The planes $\tau = -1$ and $\tau = +1$ are now invariant and carry the asymptotic flows for (2.4) as $x \to \pm\infty$. In both planes this is the system

$$p' = y, \qquad s' = -w, \qquad y' = -\beta p + \lambda s, \qquad w' = \beta s + \lambda p, \tag{2.10}$$

which is a constant coefficient linear system. Call the matrix of the right-hand side B. A simple calculation shows that the eigenvalues of B are $\pm(-\beta \pm i\lambda)^{1/2}$. Since $\beta < 0$ and $\lambda \in \mathbb{R}$, it follows that there is a two-dimensional stable subspace and a two-dimensional unstable subspace. The stable subspace (call it S) is spanned by

$$(1, 0, -\delta, -\eta) \qquad \text{and} \qquad (0, 1, -\eta, \delta), \tag{2.11}$$

where $\delta + i\eta = (-\beta + i\lambda)^{1/2}$. The unstable subspace (U) is spanned by

$$(1, 0, \delta, \eta) \qquad \text{and} \qquad (0, 1, \eta, -\delta). \tag{2.12}$$

Any solutions of (2.4) that decay as $x \to -\infty$ correspond to solutions of (2.7) that tend to $(0, -1) \in \mathbb{R}^4 \times (-1, +1]$ as $x \to -\infty$. These all lie on the global unstable manifold of this critical point (call this W^u_-). W^u_- is easily seen to be three-dimensional.

The boundary condition at $+\infty$ is satisfied by solutions on the stable manifold W^s_+ of $(0, +1)$. The following lemma then characterizes eigenvalues in terms of these manifolds.

LEMMA 2.3. *If* $W^u_- \cap W^s_+ \neq \{0\} \times (-1, +1)$, *then* λ *is an eigenvalue of* N.

Proof. Since $\{0\} \times (-1, +1) \subset W^u_- \cap W^s_+$, $W^u_- \cap \{\tau = +1\} = \emptyset$ and $W^s_+ \cap \{\tau = -1\} = \emptyset$, it follows that if $W^u_- \cap W^s_- \neq \{0\} \times (-1, +1)$, then there is a non-zero P_0 with $(P_0, \tau_0) \in W^u_- \cap W^s_+$ and $\tau_0 \neq \pm 1$. (2.5) then has a solution which decays to 0 as $x \to \pm\infty$. Since there are stable and unstable manifolds associated to eigenvalues that lie off the imaginary axis, the decay is exponential and lemma 2.1 applies.

We claim that $W^u_- \cap \{\tau = \tau_0\}$ is a two-dimensional subspace. Since the equation is linear, it must be a subspace. Near $\tau = -1$, W^u_- must be three-dimensional. Since $(0, 1)$ is an unstable eigenvector, $W^u_- \cap \{\tau = \tau_0\}$ is a two-dimensional subspace near $\tau = -1$. By solving the τ equation, there is always a tangent vector to W^u_- in the τ-direction. It follows that $W^u_- \cap \{\tau = \tau_0\}$ is two-dimensional for all $\tau_0 \neq +1$. Similar considerations hold for W^s_+.

We shall view W^u_- as a curve in the space of two-dimensional subspaces of $\mathbb{R}^4$. In order to set this up, consider firstly the second exterior power of $\mathbb{R}^4$: $\Lambda^2(\mathbb{R}^4)$. Any linear map A on $\mathbb{R}^4$ induces a corresponding $A^{(2)}$ on $\Lambda^2(\mathbb{R}^4)$ such that

$$A^{(2)}(P_1 \wedge P_2) = AP_1 \wedge P_2 + P_1 \wedge AP_2,$$

where $P_1, P_2 \in \mathbb{R}^4$ [5]. If $P_1(x)$ and $P_2(x)$ satisfy (2.5), it follows that $Q(x) = P_1(x) \wedge P_2(x)$ satisfies

$$Q' = A^{(2)}(\lambda, x)Q. \tag{2.13}$$

Choose a basis $\{e_1, e_2, e_3, e_4\}$ on $\mathbb{R}^4$; then the set $\{e_i \wedge e_j : i, j = 1, \ldots, 4\}$ forms a basis of $\Lambda^2(\mathbb{R}^4)$ and $\Lambda^2(\mathbb{R}^4) \cong \mathbb{R}^6$. A point $q \in \Lambda^2(\mathbb{R}^4)$ can then be written $Q = \sum q_{ij} e_i \wedge e_j$. (2.13) will satisfy the same conditions as (2.5) and hence can be compactified:

$$Q' = A^{(2)}(\lambda, \tau)Q, \qquad \tau' = \kappa(1 - \tau^2); \tag{2.14}$$

with an abuse of notation we shall not use the c on A to denote the change of independent variable. This gives a flow on $\mathbb{R}^6 \times [-1, +1]$.

Since (2.13) is linear, it can be projectivized to an equation on $\mathbb{R}P^5$, five-dimensional real projective space. Write this as

$$\phi' = a(\lambda, x, \phi) \tag{2.15}$$

or in compactified form as

$$\phi' = a(\lambda, x, \phi), \qquad \tau' = \kappa(1-\tau^2). \tag{2.16}$$

This last equation induces a flow on $\mathbb{R}P^5 \times [-1, +1]$. By checking in local coordinates, one sees that a is at least C^1.

If $y_1, y_2 \in \mathbb{R}^4$, then $y_1 \wedge y_2 \in \Lambda^2(\mathbb{R}^4)$ and the plane spanned by $\{y_1, y_2\}$ is associated to $y_1 \wedge y_2$. However, not all $\omega \in \Lambda^2(\mathbb{R}^4)$ can be matched to a plane in this way. An element $\omega \in \Lambda^2\mathbb{R}^4$ is said to be decomposable if $\omega = y_1 \wedge y_2$ for some $y_1, y_2 \in \mathbb{R}^4$. It is true that ω is decomposable if and only if $\omega \wedge \omega = 0$ (see [**10**]). If $Q = \sum q_{ij} e_i \wedge e_j$, this condition can be written as

$$q_{12}q_{34} - q_{13}q_{24} + q_{14}q_{23} = 0. \tag{2.17}$$

The quadric defined by (2.17) in $\mathbb{R}^6$ is invariant and hence its image in $\mathbb{R}P^5$ is also invariant; this is the Grassmannian of two-dimensional planes in $\mathbb{R}^4$: $G_{2,4}$. A flow on $G_{2,4} \times [-1, +1]$ is thus obtained.

Now set

$$Z(\lambda, x) = W_-^u \cap \{\tau = \tau(x)\},$$

where $\tau(x)$ is chosen in the usual fashion. Also set

$$\Phi(\lambda, x) = \sigma(Z(\lambda, x)),$$

where $\sigma : \mathbb{R}^4 \times [-1, +1] \to \mathbb{R}^4$ is the natural projection. It follows that $Z(\lambda, x) = (\Phi(\lambda, x)\tau(x))$. Let $P_1(\lambda, x)$ and $P_2(\lambda, x)$ be two solutions that span $\Phi(\lambda, x)$ and set $Q(\lambda, x) = P_1(\lambda, x) \wedge P_2(\lambda, x)$. $Q(\lambda, x)$ is now a curve of one-dimensional subspaces in $\Lambda^2(\mathbb{R}^4)$. If $\Pi : \mathbb{R}^6 \backslash \{0\} \to \mathbb{R}P^5$ is the natural map denoted by $\Pi(Q) = \tilde{Q}$, set $\phi(\lambda, x) = \tilde{Q}(\lambda, x)$. The curve $\zeta(\lambda, x) = (\phi(\lambda, x), \tau(x))$ is in $G_{2,4} \times [-1, +1]$.

Let $\Sigma_1 = W_+^s \cap \{\tau = +1\}$, the stable subspace of 0 inside $\tau = +1$. Set

$$\mathscr{D}_1 = \{\psi \in G_{2,4}\text{: the subspace } \Psi \text{ determined by } \psi \text{ satisfies } \Psi \cap \sigma(\Sigma_1) \neq \{0\}\}.$$

The following lemma provides the characterization we need for an eigenvalue of N. The notation ω means the omega limit set under the flow [**9**].

LEMMA 2.4. *If* $\omega(\phi(\lambda, x_0), \tau(x_0)) \cap \mathscr{D}_1 \times \{1\} \neq \varnothing$, *then* λ *is an eigenvalue of* N.

Proof. It suffices to show that

$$\omega(\phi(\lambda, x_0), \tau(x_0)) \cap \mathscr{D}_1 \times \{1\} \neq \varnothing$$

implies that $W_-^u \cap W_+^s \neq \{0\} \times (-1, +1)$, by lemma 2.3. Suppose that $W_-^u \cap W_+^s$ is trivial. With $Y(\lambda, x) = W_+^s \cap \{\tau = \tau(x)\}$ and $\Psi(\lambda, x) = \sigma(Y(\lambda, x))$ it follows that

$$\Phi(\lambda, x) \cap \Psi(\lambda, x) = \{0\}$$

for all x.

The manifold W_+^s is constructed locally in a neighbourhood of $(0, +1)$ and is attracting in that neighbourhood. Let V be such a neighbourhood and $r(x) \in Z(\lambda, x)$ with x large enough so that $\alpha r(x) \in V$ for some $\alpha > 0$. The positive number α will depend on x, but by the attractivity property of W^u_+, $\alpha r(x)$ will approach W^u_+ as $x \to +\infty$. It follows that $\rho(x) = \Pi(\text{span}\{v(x)\})$, where $\Pi : \mathbb{R}^4 \backslash \{0\} \to \mathbb{R}P^3$ is the natural projection, will approach $\Pi(W^u_+) \subset \mathbb{R}P^3$ under the flow inherited from (2.8) by projectivizing. Let U be a neighbourhood of $\Pi(W^u_+) \subset \mathbb{R}P^3$ disjoint from $\Pi(\Sigma_1)$ (each of these are closed circles in $\mathbb{R}P^3$, which is normal). It follows that $\Pi(\Phi(\lambda, x))$ intersects U if x is large enough. However, if, in $G_{2,4}$, $\phi(\lambda, x)$ lies in a neighbourhood of $\mathscr{D}_1$ for any large x, then at such a value of x, $\Pi(\Phi(\lambda, x))$ is disjoint from U, which contradicts the above.

Plücker coordinates can be used to actually compute equations on $G_{2,4}$ when viewed as a projective variety given by (2.17). This will be useful in proving the following lemma. A repeller is a set which is an attractor in backward time. An attractor is the omega limit set of a neighbourhood of itself.

LEMMA 2.5. *For each $\lambda \in \mathbb{R}$ the set $\mathscr{D}_1$ is a repeller in the flow on $G_{2,4} \times \{+1\}$.*

Proof. Change coordinates in $\mathbb{R}^4$ so that the matrix B (see (2.10)) on $\{\tau = +1\}$ has the form

$$\left(\begin{array}{cc|cc} -\delta & -\eta & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ \eta & -\delta & & \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{0}} & \delta & -\eta \\ & & \eta & \delta \end{array}\right). \tag{2.18}$$

The stable subspace is then given by the condition $y = w = 0$. In Plücker coordinates the condition defining $\mathscr{D}_1$ is $q_{34} = 0$. To see this, write out the determinant condition that two vectors spanning $\Psi \in \mathscr{D}_1$ are linearly dependent on $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$. The equation for q_{34} can be easily computed to be

$$q'_{34} = 2\delta_{34}.$$

The set $q_{34} = 0$ is thus a repeller.

The equations (2.4) (or (2.5)) are Hamiltonian and thus preserve Lagrangian planes. In the following, J detnoes the usual symplectic matrix

$$\begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix},$$

where I is the 2×2 identity.

Definition. A Lagrangian plane Ψ is a two-dimensional subspace of $\mathbb{R}^4$ that satisfies $(P_1, JP_2) = 0$ for all $P_1, P_2 \in \Psi$, where the parentheses denote the usual inner product.

The set of Lagrangian planes is denoted $\Lambda(2)$. The following lemma gives a characterization in terms of Plücker coordinates.

LEMMA 2.6. *$\Lambda(2) \subset G_{2,4}$ is given by the condition $q_{13} + q_{24} = 0$.*

Proof. It is trivial that $(P, JP) = 0$ for any vector P. It suffices to show that if P_1, P_2 span a plane Φ, then $(P_1, JP_2) = 0$ if and only if $q_{13} + q_{24} = 0$. In coordinates it is a trivial calculation to check that

$$q_{13} + q_{24} = (P_1, JP_2),$$

and the lemma follows.

To make use of $\Lambda(2)$, it is important that the trajectories of interest lie in this submanifold of $G_{2,4}$.

LEMMA 2.7. *$\phi(\lambda, x) \in \Lambda(2)$ for all λ and x.*

Proof. As mentioned above, the system (2.5) is Hamiltonian and hence preserves $\Lambda(2)\times[-1,+1]$. The curve $(\phi(\lambda, x), \tau(x))$ is the unstable manifold of the point ζ_- in $G_{2,4}\times\{-1\}$ associated to $W^u_- \cap \{\tau=-1\}$. Now $\Lambda(2)\times[-1,+1]$ is invariant and ζ_- is Lagrangian. Moreover, it has an unstable manifold in this space which is non-trivial because it comes from the τ-direction, and therefore it must be this curve $(\phi(\lambda, x), \tau(x))$.

3. *Main lemma*

The space of Lagrangian two-dimensional planes in $\mathbb{R}^4$ can be represented as a homogeneous space: $\Lambda(2) \cong U(2)/O(2)$, where $U(2)$ is the group of 2×2 unitary matrices and $O(2)$ is the subgroup of real orthogonal matrices (see [**2**] and [**8**]). An identification of the above form holds in general, but in this special low-dimensional case a very concrete visualization can be given. The space of Lagrangian 2-planes is a fibre bundle over S^1 with fibre S^2 and clutching function the antipodal map. In other words, consider $S^2\times[-1,+1]$ and identify the fibres over ± 1 by the antipodal map. In a more recent paper this is shown by Arnol'd [**3**]. A new characterization of $\Lambda(2)$ in which this can be seen very explicitly is given in the Appendix. The projection onto the circle S^1 gives the winding that is the Maslov index. On $U(2)/O(2)$ this is achieved by the map: $\det^2$.

Let $\psi \in \Lambda(2)$; Arnol'd [**3**] introduces the notion of the train of ψ, which is equivalent to the following.

Definition. The train of ψ, denoted $\mathcal{D}(\psi)$, is the set of all $\phi \in \Lambda(2)$ so that (as two-dimensional subspaces of $\mathbb{R}^4$) $\phi \cap \psi \neq \{0\}$.

Since $\Lambda(2)$ is a homogeneous space, every train $\mathcal{D}(\psi)$ is topologically identical. In the fibre bundle it covers S^1 and has the form of a sphere with its poles identified.

It is natural to use the covering space of $\Lambda(2)$ in order to exploit the winding. The covering space is $S^2\times\mathbb{R}$ as shown in the Appendix. We shall denote this covering space $C(2)$ and indicate a lift by $\hat{}$. For any $\psi \in \Lambda(2)$, its train is covered by $\hat{\mathcal{D}}(\phi)$, which can be visualized as the union of infinitely many cones (see figure 1 and [**3**]). In figure 1 each vertical slice is a disc and the fibre S^2 is formed by identifying the boundary to a point. The key point is that $\hat{\mathcal{D}}(\phi)$ divides $C(2)$ into infinitely many components [**3**].

By the lifting property of covering spaces, the curve described earlier, $\phi(\lambda, x)$, can be lifted to $C(2)$, modulo a choice of the lift of its left endpoint. Indeed a flow

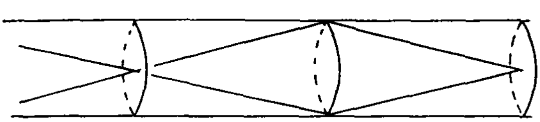

FIGURE 1. $\hat{\mathcal{D}}(\psi)$ sits inside $C(2)$.

on $C(2)\times[-1,+1]$ can be induced from that on $\Lambda(2)\times[-1,+1]$. This flow is continuous in $\lambda\in\mathbb{R}$; we can therefore put these parametrized flows together to form a flow on $C(2)\times[-1,+1]\times\mathbb{R}$.

Recall that the asymptotic system for the equation on $\mathbb{R}^4$ is given by (2.14) and is the same for $x\to\pm\infty$. Let $U=U(\lambda)$ be the unstable subspace; this is spanned by (2.12). Let $S=S(\lambda)$ be the stable subspace; it is spanned by (2.11). It is clear that both $S, U\in\Lambda(2)$. In the homogeneous space representation $\Lambda(2)\cong U(2)/O(2)$ and a subspace is represented by a unitary matrix that takes the real plane to the plane under construction. Here we are identifying $\mathbb{R}^4$ with $\mathbb{C}^2$ by (u_1, u_2, v_1, v_2) becoming (u_1+iv_1, u_2+iv_2). It follows that S is represented by

$$\frac{1}{1+\delta^2+\eta^2}\begin{bmatrix}1-i\delta & -i\eta\\ -i\eta & 1+i\delta\end{bmatrix}$$

and U is represented by

$$\frac{1}{1+\delta^2+\eta^2}\begin{bmatrix}1+i\delta & i\eta\\ i\eta & 1-i\delta\end{bmatrix}.$$

Recall that $\delta+i\eta=(-\beta+i\lambda)^{1/2}$, where the square root with branch on the negative real axis is taken. Noting that $\beta<0$, the matrices representing S and U are easily seen to be continuous in $\lambda\in\mathbb{R}$.

Now choose continuous lifts of U and S, say $\hat{U}=\hat{U}(\lambda)$ and $\hat{S}=\hat{S}(\lambda)$. Also set $\hat{\mathscr{D}}=\hat{\mathscr{D}}(S(\lambda))$, the lift of the train of S.

Fix $\lambda_1<\lambda_2$ and put together the parametrized flows on $C(2)\times[-1,+1]\times[\lambda_1,\lambda_2]$. Since $S(\lambda)$ is continuous in λ, it is easy to see that

$$\mathscr{A}_{1,2}=C(2)\times\{+1\}\times[\lambda_1,\lambda_2]\Big\backslash\bigcup_{\lambda\in[\lambda_1,\lambda_2]}(\hat{\mathscr{D}}(S(\lambda)),1,\lambda) \tag{3.1}$$

has infinitely many components. Call one such component

$$A_{1,2}=\bigcup_{\lambda}(A(\lambda),1,\lambda). \tag{3.2}$$

The main lemma is the idea of a shooting argument in the covering space. Recall that the trajectory of interest in $\Lambda(2)\times[-1,+1]$ is denoted $\zeta(\lambda,x)=(\phi(\lambda,x),\tau(x))$. Fix x_0 and choose a lift so that $\hat{\phi}(\lambda,x_0)$ varies continuously in x.

LEMMA 3.1. *If there is a component of* $\mathscr{A}_{1,2}$, *say* $A_{1,2}$ (*as in* (3.2)), *so that*

(1) $\omega(\hat{\phi}(\lambda_1,x_0),\tau(x_0))\cap[C(2)\backslash \mathrm{cl}\,(A(\lambda_1))]\times\{+1\}\neq\varnothing$,

(2) $\omega(\hat{\phi}(\lambda_2,x_0),\tau(x_0))\cap A(\lambda_2)\times\{+1\}\neq\varnothing$,

then there is a $\lambda\in[\lambda_1,\lambda_2]$ *that is an eigenvalue of* N.

Proof. Consider the set

$$X_0=\bigcup_{\lambda\in[\lambda_1,\lambda_2]}(\hat{\phi}(\lambda,x_0),\tau(x_0),\lambda)$$

and apply the flow on $C(2)\times\{+1\}\times[\lambda_1,\lambda_2]$ to this. The ω-limit set $\omega(X_0)$ is connected and by (1) and (2) above it must intersect $\partial A_{1,2}\subset\bigcup_\lambda(\mathscr{D}(S(\lambda)),\lambda)$. We want to conclude from this that there is a $\lambda\in[\lambda_1,\lambda_2]$ so that

$$\omega(\hat{\phi}(\lambda,x_0),\tau(x_0))\cap\partial A(\lambda)\times\{+1\}\neq\varnothing. \tag{3.3}$$

This, however, does not follow immediately, since $\omega(X_0)$ is not necessarily the union of

$$\omega(\hat{\phi}(\lambda, x_0), \tau(x_0), \lambda)$$

as λ varies over $[\lambda_1, \lambda_2]$.

From lemma 2.5 the set $\mathscr{D}_1$ is a repeller in $G_{2,4}\times\{+1\}$. It follows that $\mathscr{D}(S(\lambda))$ is a repeller in $\Lambda(2)\times\{+1\}$ and hence $\hat{\mathscr{D}}(S(\lambda))$ is a repeller in $C(2)\times\{+1\}$. Moreover $D=\bigcup_{\lambda\in[\lambda_1,\lambda_2]}(\hat{\mathscr{D}}(S(\lambda)), \lambda)$ is a repeller in $C(2)\times\{+1\}\times[\lambda_1, \lambda_2]$. Let Γ be the complementary attractor of D; in other words, if U is a neighbourhood of D, $\Gamma=\omega(C(2)\times\{+1\}\times[\lambda_1, \lambda_2]\backslash U)$. The complementary attractor is shown to exist in [9, p. 32] for the compact case. It can be achieved here firstly in $\Lambda(2)$ and then by lifting to $C(2)$.

Suppose that for all $\hat{\lambda}\in[\lambda_1, \lambda_2]$

$$\omega(\hat{\phi}(\hat{\lambda}, x_0), \tau(x_0), \hat{\lambda})\cap D=\varnothing;$$

it follows that it must lie in the complementary attractor so that

$$\omega(\hat{\phi}(\hat{\lambda}, x_0), \tau(x_0), \hat{\lambda})\subset\Gamma.$$

But Γ is an attractor and hence there exists $\varepsilon>0$ so that

$$\omega\left(\bigcup_{\lambda\in[\hat{\lambda}-\varepsilon,\hat{\lambda}+\varepsilon]}(\hat{\phi}(\hat{\lambda}, x_0), \tau(x_0), \hat{\lambda})\right)\subset\Gamma. \tag{3.4}$$

Since $\Gamma\subset\mathscr{A}_{1,2}$, it divides into infinitely many components. By compactness of the interval $[\lambda_1, \lambda_2]$ it follows that the quantity in (3.4) must lie in the same component for all $\hat{\lambda}$, but this contradicts the hypotheses of the lemma.

Remark. If the following hold:

(1) $$\omega(\hat{\phi}(\hat{\lambda}, x_0), \tau(x_0))\subset C(2)\backslash \mathrm{cl}\,(A_{1,2}(\lambda_1))\times\{+1\},$$

(2) $$\omega(\hat{\phi}(\hat{\lambda}, x_0), \tau(x_0))\subset A(\lambda_2)\times\{+1\},$$

then the eigenvalue $\lambda\in(\lambda_1, \lambda_2)$; in other words it is not λ_1 or λ_2. This follows easily from the above proof.

4. *Proof of theorem* 1

The strategy of the proof is to apply lemma 3.1 with $\lambda_1=0$ and $\lambda_2\gg 1$. We have for each $\lambda\in\mathbb{R}$ a distinguished trajectory for the flow on $C(2)\times[-1, +1]$, namely $\hat{\zeta}(\lambda, x)=(\hat{\phi}(\lambda, x), \tau(x))$. This trajectory satisfies the appropriate boundary condition as $x\to-\infty$, namely

$$\hat{\phi}(\lambda, x)\to\hat{U}(\lambda)$$

as $x\to-\infty$, where $\hat{U}(\lambda)$ is some lift of the unstable subspace $U(\lambda)$ chosen continuously in $\lambda\in\mathbb{R}$, as in § 3. Since $U(\lambda)\notin\mathscr{D}(S(\lambda))$ for all $\lambda\in\mathbb{R}$, we can choose a component of $\mathscr{A}_{1,2}$ as in (3.2), say $A_{1,2}$, so that

$$\hat{U}(\lambda)\in A_{1,2}$$

for all $\lambda\in\mathbb{R}$. We shall now analyse the limit cases and show that the hypotheses of lemma 3.1 are satisfied under the condition $P-Q\neq 0, 1$.

The first step is to analyse the behaviour of $\zeta(\lambda, x)$ for $\lambda \gg 1$. Consider system (2.5) again,

$$P' = A(\lambda, x)P,$$

and rescale by introducing $\xi = \lambda^{-1/2}x$ and

$$\tilde{y} = \lambda^{-1/2}y, \qquad \tilde{w} = \lambda^{-1/2}w. \tag{4.1}$$

This leads to the system

$$\dot{p} = \tilde{y}, \qquad \dot{s} = -\tilde{w},$$
$$\dot{\tilde{y}} = -(g(x)/\lambda^{1/2})p + s, \qquad \dot{\tilde{w}} = (h(x)/\lambda^{1/2})s + p, \qquad \dot{\tau} = (\kappa/\lambda^{1/2})(1-\tau^2), \tag{4.2}$$

where $\dot{} = d/d\xi$, when compactified with τ. Abbreviate this as

$$\dot{\tilde{P}} = A(\lambda, \tau)\tilde{P}, \qquad \dot{\tau} = (\kappa/\lambda^{1/2})(1-\tau^2). \tag{4.3}$$

Now let $\lambda \to +\infty$; the limiting system of (4.2) is

$$\dot{p} = \tilde{y}, \qquad \dot{s} = -\tilde{w}, \qquad \dot{\tilde{y}} = s, \qquad \dot{\tilde{w}} = p, \qquad \dot{\tau} = 0. \tag{4.4}$$

This equation is linear and autonomous. Each τ = constant slice is invariant. The eigenvalues of the matrix for (4.4) are $\pm(\pm i)^{1/2}$. There is therefore a two-dimensional stable and a two-dimensional unstable subspace in each τ-slice. The equation (4.2) can be transformed and projectivized as before. This produces a transformed flow on $\Lambda(2)\times[-1,+1]$, indexed by $\lambda \in \mathbb{R}$. Let $\hat{\zeta}(\lambda, x)$ be the image of $\zeta(\lambda, x)$ under this transformation. As $x \to -\infty$, $\hat{\zeta}(\lambda, x) \to \hat{\zeta}_-$, the critical point associated to the unstable subspace.

Now if $\lambda \to +\infty$, in the limit system on $\Lambda(2)\times[-1,+1]$, $\hat{\zeta}_- \times[-1,+1]$ is a curve of critical points. Since $\hat{\zeta}_-$ came from the unstable subspace, each one is attracting. Let $N\times[-1,+1]$ be an attracting neighbourhood of $\hat{\zeta}_-\times[-1,+1]$. If $\lambda \gg 1$, $N\times[-1,+1]$ remains an attracting neighbourhood and therefore $\hat{\zeta}(\lambda, x)$ lies in this set for all x. It follows that $\hat{\phi}(\lambda, x)$ lies in a neighbourhood in $C(2)$ of its value as $x \to -\infty$, since the transformation is independent of τ. This proves the following lemma.

LEMMA 4.1. *If* $\lambda \gg 1$,

$$\omega(\hat{\phi}(\lambda, x_0), \tau(x_0)) \subset A(\lambda)\times\{+1\}.$$

Setting λ_2 to be this value of λ, part (2) of lemma 3.1 is seen to be satisfied.

Now consider the case $\lambda = 0$. The original linear equations (2.1) simplify to

$$L_+p = 0, \qquad L_-q = 0, \tag{4.5}$$

which can be written

$$p'' + g(x)p = 0, \qquad q'' + h(x)q = 0. \tag{4.6}$$

Each of these leads to a flow on $\mathbb{R}^2$ which can be projectivized. This renders a flow on $\mathbb{R}P^1\times\mathbb{R}P^1$, which is a torus. We claim that this torus lies in $\Lambda(2)$ in a natural way and is invariant when $\lambda = 0$.

Let U be a unitary matrix; viewing $\Lambda(2) \cong U(2)/O(2)$, a solution on $\Lambda(2)\times[-1,+1]$ is given by $(U(x), \tau(x))$, where $U(x)$ is a matrix that takes Σ_R (the real plane) to the plane $\Phi(0, x)$. Since equation (2.5) uncouples, U can be taken to be a diagonal matrix. Σ_R is spanned by $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$ and $\Phi(0, x)$ is spanned by $(p, 0, y, 0)$ and $(0, s, 0, w)$, where $y = p'$ and p satisfies

$$L_+ p = 0,$$

which decays as $x \to -\infty$, $w = -s'$ and s satisfies

$$L_- s = 0,$$

also decaying as $x \to -\infty$.

In $\mathbb{C}^2$ this is the complex line determined by applying

$$\begin{pmatrix} p + iy & 0 \\ 0 & s + iw \end{pmatrix} \tag{4.7}$$

to Σ_R. Normalizing, (4.7) becomes

$$\begin{pmatrix} \dfrac{p + iy}{\sqrt{p^2 + y^2}} & 0 \\ 0 & \dfrac{s + iw}{\sqrt{s^2 + w^2}} \end{pmatrix}. \tag{4.8}$$

The matrix (4.8) is $U(x)$. The correspondence $U(2)/O(2) \cong S(2)$ (see the Appendix) is given by $S = UU^{\mathrm{T}}$. Therefore the matrix in $S(2)$ corresponding to $\zeta(0, x)$ is

$$\begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{pmatrix}, \tag{4.9}$$

where

$$\theta_1 = 2 \tan^{-1} (y/p), \qquad \theta_2 = 2 \tan^{-1} (w/s).$$

Recall that $s = -q$ and so

$$\theta_2 = -2 \tan^{-1} (w/q) = -2 \tan^{-1} (q'/q).$$

The set of matrices under consideration is therefore the diagonal matrices in $S(2)$, which, as commented in the Appendix, form a torus $T \subset \Lambda(2) \cong S(2)$. The plane $(\theta_1, \theta_2) \in \mathbb{R}^2$ is the covering space of T. To see how it relates to the covering space of $\Lambda(2)$, let $\rho : C(2) \to \Lambda(2)$ be the covering. The set $\rho^{-1}(T)$ can be obtained from $\mathbb{R}^2$ by identifying the lines $\theta_2 = \theta_1 \pm 2\pi$ in the obvious way.

Consider the behaviour of $(\theta_1(x), \theta_2(x))$ as $x \to -\infty$, where $\zeta(0, x) = (\phi(0, x), \tau(x))$ is as usual the trajectory of interest in $\Lambda(2) \times [-1, +1]$ and $\phi(0, x)$ is represented by

$$\begin{pmatrix} e^{i\theta_1(x)} & 0 \\ 0 & e^{i\theta_2(x)} \end{pmatrix}.$$

From the analysis of the uncoupled systems it is easy to check that

$$\theta_1(x) \to 2 \tan^{-1} (\sqrt{-\beta}) = \theta_1^-, \qquad \theta_2(x) \to -2 \tan^{-1} (\sqrt{-\beta}) = \theta_2^-$$

as $x \to -\infty$ (see figure 2).

https://doi.org/10.1017/S014338570000938X Published online by Cambridge University Press

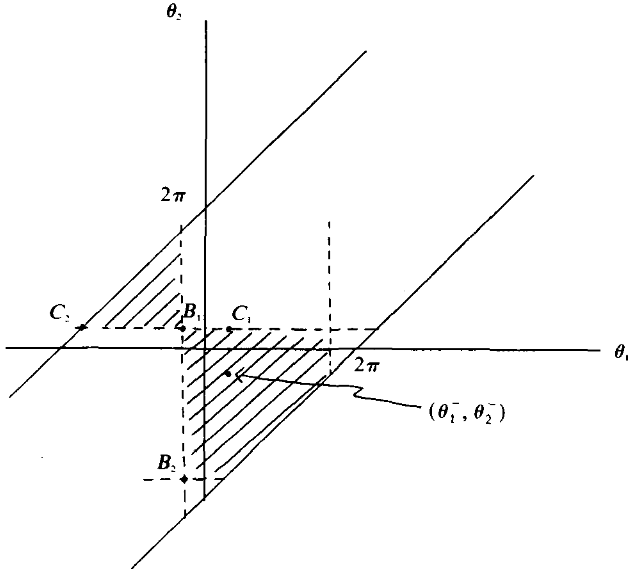

FIGURE 2

Note that in these coordinates the stable subspace in $\tau=+1$ is not the real plane but corresponds to

$$\theta_1=-2\tan^{-1}(\sqrt{-\beta}),\qquad \theta_2=2\tan^{-1}(\sqrt{-\beta}).$$

Recall that $\mathscr{D}_1$ is the set of subspaces intersecting this space. It can be easily checked that $\hat{\mathscr{D}}_1\cap T$ consists of the lines

$$\theta_1=-2\tan^{-1}(\sqrt{-\beta})+2n\pi,\qquad \theta_2=2\tan^{-1}(\sqrt{-\beta})+2m\pi,$$

where $m, n\in\mathbb{Z}$.

By lemmas 3.1 and 4.1, theorem 1 will be proved by showing that $P-Q\neq 0, 1$ implies that $(\theta_1(x), \theta_2(x))$ does not tend, as $x\to+\infty$, to any point in the component of $C(2)\backslash\hat{\mathscr{D}}_1$ that contains (θ_1^-, θ_2^-); this is the shaded region in figure 2. In the following, note that in $C(2)\supset\rho^{-1}(T)$ we can take $(\theta_1+2n\pi, \theta_2-2n\pi)$ as equivalent to (θ_1, θ_2).

Firstly calculate $\theta_1(x)$ as $x\to+\infty$. This comes from the equation $L_-q=0$. Recall that $q(x)=u(x)$ is the solution of interest. Let

$$\tilde{\theta}=2\tan^{-1}(q'/q);$$

then $\theta_2=-\tilde{\theta}$. Since $u(x)\to 0$ as $x\to\pm\infty$,

$$\tilde{\theta}(x)\to -2\tan^{-1}(\sqrt{-\beta})+2m\pi$$

as $x\to+\infty$ for some $m\in\mathbb{Z}$. It follows that

$$\theta_2(x)\to 2\tan^{-1}(\sqrt{-\beta})+2m\pi$$

as $x\to+\infty$ for some $m\in\mathbb{Z}$.

For $\theta_1(x)$ consider $L_+p=0$; there are two possibilities for $\theta_1(x)$ depending on whether $p(x)$ is bounded or not.

Case 1. $p(x)$ is bounded implies that

$$\theta_1(x) \to -2\tan^{-1}(\sqrt{-\beta}) + 2n\pi$$

for some $n \in \mathbb{Z}$ as $x \to +\infty$.

Case 2. $p(x)$ is unbounded; then

$$\theta_1 \to 2\tan^{-1}(\sqrt{-\beta}) + 2n\pi.$$

Let A be the component of $C(2)\backslash\hat{\mathscr{D}}_1 \cap T$ containing (θ_1^-, θ_2^-), i.e. the shaded region in figure 2. In case 1,

$$\begin{aligned}(\theta_1, \theta_2) &\to (2\tan^{-1}(\sqrt{-\beta}) + 2n\pi, 2\tan^{-1}(\sqrt{-\beta}) + 2m\pi)\\ &\sim (2\tan^{-1}(\sqrt{-\beta}), 2\tan^{-1}(\sqrt{-\beta}) + 2(m-n)\pi).\end{aligned}$$

The only possible values for this asymptotic value of (θ_1, θ_2) that are on $\partial A(0)$ are C_1, C_2 in figure 2. These correspond to $m - n = 0, -1$ respectively.

In case 2,

$$(\theta_1, \theta_2) \to (-2\tan^{-1}(\sqrt{-\beta}), 2\tan^{-1}(\sqrt{-\beta}) + 2(m-n)\pi).$$

Again the only possible points on $\partial A(0)$ that have this form are B_1, B_2 in figure 2 and these have $m - n = 0, -1$.

To complete the proof, one observes that by applying Sturm-Liouville theory to both L_+ and L_- we obtain

$$P = n, \qquad Q = m.$$

If $P - Q \neq 0, 1$, then $m - n \neq 0, -1$ and the asymptotic values of (θ_1, θ_2) are not on $\partial A(0)$. Therefore part (1) of lemma 3.1 is satisfied with $\lambda_1 = 0$ and it follows that there is a $\lambda \in [0, \infty)$ which is an eigenvalue. Since the stronger statements hold as in the remark following lemma 3.1, the forced eigenvalue λ must lie in $(0, \infty)$.

5. *Application*

Optical waveguides have attracted much attention recently. A case of particular interest is that of three layers of different dielectric material. The geometry is essential two-dimensional. It is assumed that the interfaces are planar and that propagation is in the plane of these interfaces. The bounding media are assumed to be non-linear with Kerr-like (cubic) response to the electromagnetic field. The sandwiched medium is linear.

With the interfaces at $x = \pm d$, the refractive index is

$$n(x, |u|^2) = \begin{cases} n_0 + \alpha_0|u|^2, & |x| > d,\\ n_1, & |x| \leq d,\end{cases}$$

where n_0, α_0 and n_1 are constants (see [1] and [15]).

We shall take $f(x, |u|^2)$ to be a smooth approximation of $n(x, |u|^2)$ so that the hypotheses (H1) and (H2) are satisfied. To be precise, pick an $\varepsilon > 0$ and let $g_\varepsilon(x)$ be a smooth function which takes the following values:

$$\begin{cases} 0 & \text{if } |x| \geq d + \varepsilon,\\ 1 & \text{if } |x| \leq d + \varepsilon.\end{cases}$$

Set

$$h_\varepsilon(x, |u|^2) = g_\varepsilon(x)n_1 + (1 - g_\varepsilon(x))(n_0 + \alpha_0|u|^2);$$

then $h_\varepsilon = n$ if $|x| \leq d - \varepsilon$ or $|x| \geq d + \varepsilon$. Moreover, the hypothesis (H1) is satisfied although the bounds become large as $\varepsilon \to 0$. As $x \to +\infty$, $h_\varepsilon(x, 0) \to n_0$ and so we set

$$f_\varepsilon(x, |u|^2) = h_\varepsilon(x, |u|^2) - n_0;$$

(H2) is then satisfied also. In fact, $f_\varepsilon(x, 0) = 0$ if $|x|$ is large enough. It follows that any standing wave $u(x)$ satisfies (H3).

The evolution equation of interest is now (1.1) with $f = f_\varepsilon$ for some $\varepsilon > 0$. To apply the theorems we have proved, we need to regularize the problem in the fashion described above. It is easiest to see how the standing waves are constructed for the limit problem, but it will be clear that these can be approximated by solutions of the regularized problem. This will lead to a family of standing waves $u^\varepsilon(x) \to u(x)$ as $\varepsilon \to 0$. Let L_+^ε and L_-^ε be the usual operators for the ε-approximate problem and

$$N^\varepsilon = \begin{pmatrix} 0 & L_-^\varepsilon \\ -L_+^\varepsilon & 0 \end{pmatrix}.$$

The solutions of $L_+^0 p = 0$ and $L_-^0 q = 0$ can be determined easily and it will be seen that these perturb to $p_\varepsilon(x)$ and $q_\varepsilon(x)$. This allows one to compute P and Q for $\varepsilon \neq 0$.

For the wave that has a certain limiting configuration $u(x)$, we shall conclude that N^ε has a real positive eigenvalue if $\varepsilon > 0$ is small enough. It is likely that N^0 could be proved to have an unstable eigenvalue by a perturbation argument [**12**]; however, we do not explore this here.

It is described in [**11**] how to construct standing waves for the $\varepsilon = 0$ problem. These are constructed by superimposing the phase portraits of the two equations

$$u'' + (n_0 + \alpha_0|u|^2)u = 0 \tag{5.1}$$

and

$$u'' + n_1 u = 0. \tag{5.2}$$

Trajectories must be found which lie on a phase curve of (5.1) on the sets $(-\infty, -d)$ and $(d, +\infty)$. They must lie on a phase curve of (5.2) for $x \in (-d, +d)$. Moreover, the following matching must hold:

$$\lim_{x \to d^-} u(x) = \lim_{x \to d^+} u(x), \qquad \lim_{x \to d^-} u'(x) = \lim_{x \to d^+} u'(x), \tag{5.3}$$

and similarly at $x = -d$.

Various configurations have been found by analytical and numerical methods and are discussed in [**11**]. We shall be interested in those that are the symmetric waves beyond the bifurcation value. These have a phase portrait where the linear part of the trajectory extends outside the homoclinic orbit of the non-linear equation (see figure 3). We shall call these type-S orbits. The application of the theorem will lead to the conclusion that type-S orbits are unstable, in the sense that the associated N^ε has a real positive eigenvalue.

An alternative representation of $\varepsilon = 0$ standing waves is given as follows. If $\beta + n_0 < 0$, (5.1) has a homoclinic orbit. Let C be the part of this orbit lying in the

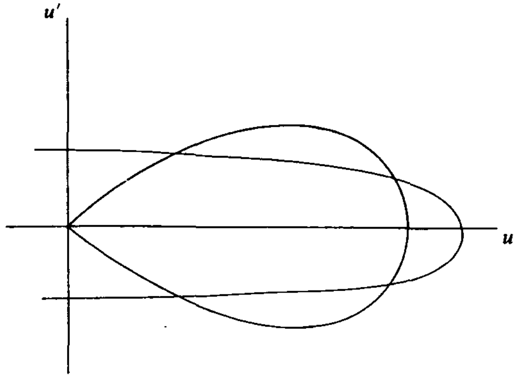

FIGURE 3. Type-S orbits.

set $\{u' \geq 0\}$ and C^- the rest. Apply the 2-D time map T of the flow associated to (5.2). Standing waves are determined by intersections of $T(C)$ and C^-. If $T(C)$ and C^- cross transversely, we shall call the wave non-degenerate. It is not hard to show that a non-degenerate wave $u(x)$ perturbs to a standing wave solution of the $\varepsilon \neq 0$ problem if ε is sufficiently small.

Let $v^-(y)$ be tangent to C^- at $y \in C^-$ pointing away from $(0, 0)$. Let $v^T(y)$ be tangent to $T(C)$ at $y \in T(C)$ pointing away from $(0, 0)$. If $y \in T(C) \cap C^-$ corresponds to an orbit of type S, then $v^- \times v^T > 0$, i.e. v^T is in a counterclockwise rotation from v^-. In fact this condition characterizes orbits of type S (see figure 4).

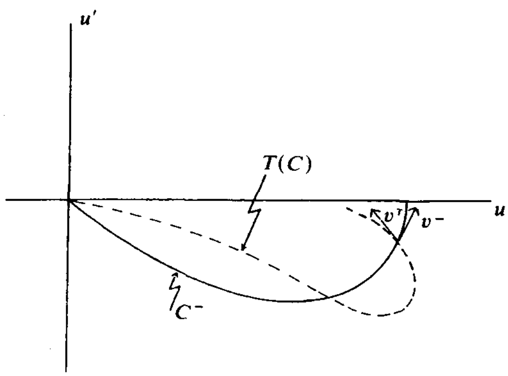

FIGURE 4. The intersection of $T(C)$ and C^-.

The next step is to compute P and Q for a wave $u_\varepsilon(x)$ ($\rightarrow u(x)$) of type S. Clearly $u_\varepsilon(x) \geq 0$ and so $Q = 0$. P is determined by considering a solution of

$$L_+^\varepsilon p = 0. \tag{5.4}$$

However, (5.4) is the equation of variations of the standing wave equations.

For $\varepsilon = 0$ a solution can easily be constructed by following a tangent vector around the orbit. On $(-\infty, -d)$ the vector will be tangent to the homoclinic orbit. For $x \in (-d, +d)$ it can be compared with the tangent vector to the solution of (5.2) and

https://doi.org/10.1017/S014338570000938X Published online by Cambridge University Press

then with the tangent vector to the homoclinic again for $x > +d$. The resulting solution will necessarily have two zeros.

If $\varepsilon > 0$ and $u(x)$ is non-degenerate, the above argument can be perturbed to produce a solution of $L_+^\varepsilon p = 0$ with two zeros. It follows that $P \geq 2$ and that $P - Q \geq 2$. By theorem 1, N^ε has a real positive eigenvalue.

Acknowledgements. For the understanding of the geometry and topology of this problem, the author owes much to Nick Ercolani. I also thank Martin Guest, Mark Levi and Dave McLaughlin for many useful conversations. Jerry Moloney and George Stegeman patiently explained the physics of optical waveguides to me and I thank them for this. This research was supported in part by the National Science Foundation NSF Grants DMS 8501961 and DMS 8507056, Air Force Office of Scientific Research AFOSR Grant #83-0027, and Army Research Office ARO Contract DAAG-29-83-K-0029.

Appendix

The space of the Lagrangian two-dimensional subspace of $\mathbb{R}^4$ is identified as the homogeneous space $U(2)/O(2)$ [**2**]. The following identification of $\mathbb{R}^4$ with $\mathbb{C}^2$ is being used here. Let (u_1, u_2, v_1, v_2) be coordinates in $\mathbb{R}^4$ and set $z = (u_1 + iv_1, u_2 + iv_2)$. $U(2)$ is then the space of 2×2 unitary matrices of $\mathbb{C}^2$ and $O(2)$, the subgroup of real orthogonal matrices.

Let $S(2)$ denote the 2×2 symmetric unitary matrices. This is a smooth submanifold of $U(2)$. $U(2)$ acts on the left on $S(2)$ by the prescription $U \in U(2)$ acts on $S \in S(2)$ to give $U^T S U$. Given any $S \in S(2)$, it can be obtained from the identity by the action of some $U \in U(2)$. One can find U such that $U^T U = S$ by just taking U to be the symmetric, unitary square root of S. It follows that $U(2)$ acts transitively on $S(2)$.

If $U \in S(2)$ lies in the isotropy subgroup at I, then $U^T U = I$; since $\bar{U}^T U = I$, we see that U must be real and hence is in $O(2)$. It therefore follows that $\Lambda(2) \cong U(2)/O(2) \cong S(2)$.

The manifold $S(2)$ can be visualized very simply. It consists of matrices $A + iB$ with A and B symmetric, while $A^2 + B^2 = I$ and $AB - BA = 0$. A and B therefore commute and are simultaneously diagonalizable by a special orthogonal matrix.

Let $R \in SO(2)$ diagonalize A and B so that $A + iB = R(D_1 + iD_2)R^T$, where

$$D_1 + iD_2 = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\phi} \end{pmatrix}.$$

The space of matrices of this diagonal form is a torus and $S(2)$ is therefore a quotient space of $T \times SO(2)$.

To see which different matrices of the above form have to be identified, firstly notice that a diagonal D commutes with a non-trivial $R \in SO(2)$ if and only if D is a scalar matrix, i.e. $D = \alpha I$ for some $\alpha \in \mathbb{C}$, $|\alpha| = 1$. If D is not scalar but diagonal, then $R_1 D R_1^T = R_2 D R_2^T$ if and only if $R_1 = \pm R_2$. If D is diagonal, then RDR^T is also diagonal when $R = \pm I$ or

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This last case of R switches the two diagonal elements. Thus each non-scalar $D \in T$ generates a circle that intersects T at the matrix with switched diagonal elements.

To see how $S(2)$ is constructed, take a fundamental domain $\mathscr{D}$ for the torus as one bounded by diagonal lines; for instance, the lines

$$\phi = -\omega + 2\pi, \qquad \phi = \omega - 2\pi, \qquad \phi = -\omega, \qquad \phi = \theta + 2\pi$$

(see figure 5).

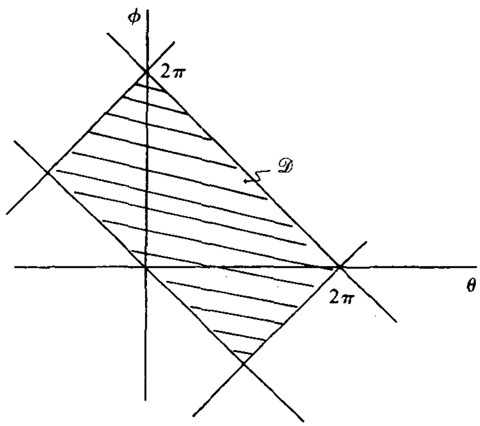

FIGURE 5. The fundamental domain $\mathscr{D}$.

Now imagine attaching a circle to each point in this region $\mathscr{D}$. Along the diagonal $\phi = \theta$, collapse the circles to a point. Further, paste the circle associated to (θ, ϕ) to that of (ϕ, θ) via the antipodal map. We now have a solid cylinder: associated to each line of the form $\phi = -\theta + c$ (c = constant) is a disc (see figure 6). The boundary of this disc is the circle associated to matrices with $\phi = \theta + 2\pi$ and $\phi = \theta - 2\pi$; these are scalar matrices and therefore this circle should be collapsed to a point. The disc now becomes a 2-sphere. The lines $\phi = -\theta + 2\pi$ and $\phi = -\theta + 2\pi$ are to be identified and so the associated spheres pasted together. On the disc the pasting is a rotation composed with an inversion; this is the antipodal map.

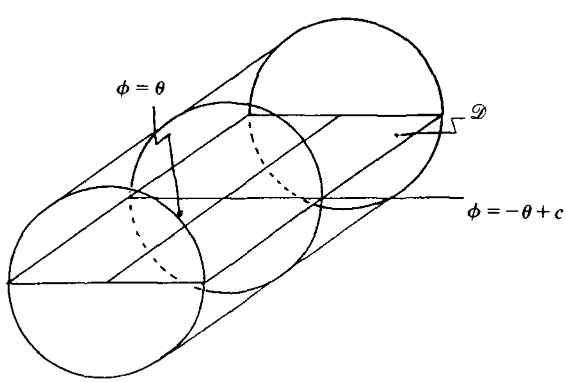

FIGURE 6. Geometric representation of $S(2)$.

https://doi.org/10.1017/S014338570000938X Published online by Cambridge University Press

REFERENCES

[1] N. N. Akhmediev. Novel class of nonlinear surface waves: asymmetric modes in a symmetric layered structure. *Sov. Phys. JETP* **56** (1982).

[2] V. I. Arnol'd. Characteristic class entering in quantization conditions. *Funct. Anal. Appl.* **1** (1967), 1-14.

[3] V. I. Arnol'd. The Sturm theorem and symplectic geometry. *Funct. Anal. Appl.* **19** (1985), 251-259.

[4] H. Berestycki & P.-L. Lions. Théorie des points critiques et instabilité des ordres stationnaires pour des équations de Schrödinger non linéaires. *C. R. Acad. Sci., Paris* **300**, série 1, #10 (1985), 319-322.

[5] R. L. Bishop & R. J. Crittenden. *Geometry of Manifolds.* Academic Press, New York (1964).

[6] R. Bott. On the iterations of closed geodesics and the Sturm intersection theory. *Commun. Pure Appl. Math.* **9** (1956), 171-206.

[7] T. Cazenave & P.-L. Lions. Orbital stability of standing waves for some nonlinear Schrödinger equations. *Commun. Math. Phys.* **85** (1982), 549-561.

[8] C. Conley. An oscillation theorem for linear systems with more than one degree of freedom. *IBM Technical Report* #18004. IBM Watson Research Center, Yorktown Heights, New York (1972).

[9] C. Conley. *Isolated Invariant Sets and the Morse Index.* CBMS Regional Conference Series in Mathematics **38**. AMS, Providence, RI (1978).

[10] Griffiths & J. Harris. *Principles of Algebraic Geometry.* Wiley, New York (1978).

[11] C. Jones & J. Moloney. Instability of nonlinear waveguide modes. *Phys. Lett. A* **117** (1986).

[12] T. Kato. *Perturbation Theory for Linear Operators.* Springer-Verlag, New York (1966).

[13] A. Newell. *Solitons in Mathematics and Physics.* SIAM, Philadelphia (1985).

[14] J. Shatah & W. Strauss. Instability of nonlinear bound states. *Commun. Math. Phys.* **100** (1985), 173-190.

[15] G. I. Stegeman & C. T. Seaton. Nonlinear surface polaritons. *Opt. Lett.* **9** (1984).

[16] W. Strauss. The nonlinear Schrödinger equation. In *Contemporary Developments in Continuous Mechanics and Partial Differential Equations*, ed. G. M. de la Penha & L. A. Medeiros. North-Holland, Amsterdam (1978).

[17] M. Weinstein. Modulational stability of ground states of nonlinear Schrödinger equations. *SIAM J. Appl. Math.* **16** (1985).