TWO-SCALE EIGENFUNCTION METHODS
WITH AN APPLICATION TO THE VISUAL SYSTEM*

By

L. SIROVICH AND B. W. KNIGHT

Rockefeller University and Brown University

1. Introduction. We investigate the eigentheory of Hermitian integral operators in unbounded domains. Although the analysis and methods apply to higher dimensions, the application lies in two dimensions and so do our examples. Our problem has its origin in studies of vision and in what follows we make use of mathematical models which arise from laboratory data. The seldom analysed features of our eigenfunction problem arise in general situations where two scales are present, as they are here. The present paper is a continuation of our earlier investigations in which the assumption of one-dimensionality was adopted [1, 2, 3]. A short general account and related ideas appear elsewhere [4, 5]. We start with a brief background to the problem.

A visual system is stimulated by an excitation $e(\mathbf{x})$, which in the present context may be considered as a pattern of illumination. This elicits a response, which in a useful range is linearly related to the excitation

$$r(\mathbf{x}) = \int K\{\mathbf{x}, \mathbf{y}\} e(\mathbf{y}) \, d\mathbf{y}. \tag{1.1}$$

From the origin of the problem the dimension of $\mathbf{x}$ and $\mathbf{y}$ is 2. It is natural to consider the eigenfunctions of the kernel $K\{\mathbf{x}, \mathbf{y}\}$ which in the present context means self-replicating patterns,

$$\int K\{\mathbf{x}, \mathbf{y}\} \psi(\mathbf{y}) \, d\mathbf{y} = \lambda \psi(\mathbf{x}). \tag{1.2}$$

Those eigenfunctions which have the largest eigenvalues play a special role and are referred to as principle eigenfunctions. With regard to the visual system these represent the patterns most strongly amplified. Assumptions based on experience specify the kernel to a form which is tractable.

In the context of vision $K\{\mathbf{x}, \mathbf{y}\}$, referred to as the point spread function [6, 7], represents the response at $\mathbf{x}$ due to a unit source at $\mathbf{y}$. Experience bears out the approximation that cause and effect are reciprocal, in the sense that the kernel is symmetric, i.e.,

$$K\{\mathbf{x}, \mathbf{y}\} = K\{\mathbf{y}, \mathbf{x}\}. \tag{1.3}$$

Without loss of generality we may write the kernel in terms of mean and relative coordinates,

$$K\{\mathbf{x}, \mathbf{y}\} = K((\mathbf{x} + \mathbf{y})/2, \mathbf{x} - \mathbf{y}) = K((\mathbf{x} + \mathbf{y})/2, \mathbf{y} - \mathbf{x}), \tag{1.4}$$

where the last is a consequence of (1.3). Considerable progress in the theory of vision is based on the assumption of homogeneity [8, 9]

$$K = K(\mathbf{x} - \mathbf{y}). \tag{1.5}$$

The resulting eigentheory is then straightforward. Eigenfunctions are sinusoids, the eigenvalues correspond to the Fourier transform of K, and the spectrum is continuous. However, (1.5) is recognized as an idealization which yields some misleading results even though the departure from homogeneity is in a sense locally slight [10, 11]. Analytically, this *slow* departure from homogeneity is represented by writing

$$K = K(\varepsilon(\mathbf{x} + \mathbf{y})/2, \mathbf{x} - \mathbf{y}) = K(\mathbf{q}, \mathbf{u}), \tag{1.6}$$

where

$$\mathbf{u} = \mathbf{x} - \mathbf{y}, \qquad \mathbf{q} = \varepsilon(\mathbf{x} + \mathbf{y})/2 \tag{1.7}$$

and ε is the small parameter measuring inhomogeneity.

Further simplification of (1.6) results from consideration of the underlying mechanisms of visual information processing. Common visual nerve networks are organized so that the excitation (due to illumination) is collected over an approximately circular area of say radius d [12, 13]. Both the amplitude of response, say A, and the collection radius d vary at a rate which is slow compared to the scale of d. A general model incorporating these inhomogeneities has the form

$$K = A(q)M(u/d(q)). \tag{1.8}$$

Note that all arguments are scalar in (1.8), so that it contains the assumption of circular symmetry of collection. In visual terms, d is referred to as resolution and A as sensitivity. Visual systems are commonly organized in an opponent manner [7] so that instead of (1.8) one has

$$K = A_1(q)M_1(u/d_1(q)) - A_2(q)M_2(u/d_2(q)). \tag{1.9}$$

This is referred to as a center-surround organization of excitation and inhibition. In what follows we consider the more general kernel,

$$K = K(q, u), \tag{1.10}$$

and illustrate the results for particular models having the form (1.9).

2. Generalities. We now briefly develop the general procedure for dealing with the eigenfunction problem (see also refs. 4, 5)

$$\int K(\varepsilon(\mathbf{x} + \mathbf{y})/2, \mathbf{x} - \mathbf{y})\psi(\mathbf{y})\, d\mathbf{y} = \lambda\psi(\mathbf{x}), \tag{2.1}$$

where ε is a parameter of smallness. In order to find a solution we express the eigenfunction in the form

$$\psi = \exp[i\phi(\mathbf{q};\varepsilon)/\varepsilon], \tag{2.2}$$

where in the present instance

$$\mathbf{q} = \varepsilon\mathbf{x}. \tag{2.3}$$

If we set $\mathbf{x} - \mathbf{y} = \mathbf{u}$, (2.1) may be arranged in the form

$$\int K(\mathbf{q} - \varepsilon\mathbf{u}/2, \mathbf{u})\exp[i(\phi(\mathbf{q} - \varepsilon\mathbf{u}) - \phi(\mathbf{q}))/\varepsilon]\, d\mathbf{u} = \lambda. \tag{2.4}$$

Thus in the limit $\varepsilon \downarrow 0$,

$$\tilde{K}(\mathbf{q}, \partial\phi_0/\partial\mathbf{q}) = \lambda, \tag{2.5}$$

where $\tilde{K}$ denotes the Wigner transform [14]

$$\tilde{K}(\mathbf{q}, \mathbf{p}) = \int K(\mathbf{q}, \mathbf{u})\exp[-i\mathbf{p} \cdot \mathbf{u}]\, d\mathbf{u}, \tag{2.6}$$

and

$$\phi_0(\mathbf{q}) = \phi(\mathbf{q}; 0). \tag{2.7}$$

We write

$$\mathbf{p} = \frac{\partial\phi_0}{\partial\mathbf{q}}, \tag{2.8}$$

and then by standard methods the first-order partial differential equation (2.5) is reduced to the Hamiltonian system [15]

$$\frac{d\mathbf{q}}{dt} = \frac{\partial\tilde{K}}{\partial\mathbf{p}}, \qquad \frac{d\mathbf{p}}{dt} = -\frac{\partial\tilde{K}}{\partial\mathbf{q}} \tag{2.9}$$

while phase ϕ_0 is determined by

$$\frac{d\phi_0}{dt} = \mathbf{p} \cdot \frac{d\mathbf{q}}{dt}. \tag{2.10}$$

The *time* variable in the above simply denotes the natural variable along an orbit. $\tilde{K}$, which plays the role of a Hamiltonian, is an invariant of (2.9).

These steps form the first stage in a formal perturbation procedure based on the expansion,

$$\phi(\mathbf{q}, \varepsilon) = \phi_0(\mathbf{q}) + \varepsilon\phi_1(\mathbf{q}) + \cdots. \tag{2.11}$$

It proves convenient to write ϕ_1 as

$$\phi_1 = -i \ln A(\mathbf{q}),$$

so that

$$\psi = A(\mathbf{q})(1 + O(\varepsilon))\exp\left[i\phi_0(\mathbf{q})/\varepsilon\right]. \tag{2.12}$$

As may be verified, the next order yields the amplitude equation,

$$\frac{\partial}{\partial \mathbf{q}} \cdot \left(A^2 \frac{\partial \tilde{K}}{\partial \mathbf{p}} \right) = 0. \tag{2.13}$$

The above development must be applied to each branch of equation (2.5). Thus if a subscript is used to denote each such branch the full solution has the form

$$\psi = \sum_j \psi_j \sim \sum_j A_j \exp\left[i\phi_j/\varepsilon\right]. \tag{2.14}$$

(We have dropped the now superfluous zero subscript on the phase ϕ.)

The steps just outlined parallel those associated with the WKB method [16, 17] (or the semi-classical approximation), but now appear in a broader context. In fact the only restriction on the kernel of (2.1) is that the integral exist in some sense. For example, the case of pseudo-differential operators is included. At this point the classic development of Keller [18, 19] can be carried over directly. Each WKB solution breaks down when A diverges, e.g., at a caustic or a focus. The analysis must then be reexamined and the resulting local analysis results in connection formulas. The Keller procedure is therefore seen to apply to the more general case just discussed. These remarks are only known to apply when K represents an *integrable Hamiltonian*. For further remarks in this vein see reference [4].

3. The separable case. The kernels which arise out of the vision problems discussed in the Introduction are well approximated by kernels of the general form

$$K = K(q, u), \tag{3.1}$$

for which the Wigner transform is

$$\tilde{K} = \tilde{K}(q, p), \tag{3.2}$$

with

$$p^2 = p_1^2 + p_2^2, \qquad q^2 = q_1^2 + q_2^2. \tag{3.3}$$

Thus the *Hamiltonian* $\tilde{K}$ is separable as it is for example for the central force problem of mechanics [20].

In order to treat a Hamiltonian of the form (3.2) we introduce cylindrical coordinates (q, θ) via a canonical transformation [20]. Hence

$$p^2 = \rho^2 + \mu^2/q^2, \tag{3.4}$$

where ρ is the radial and μ/q the tangential component of *momentum*. In the present context

$$\rho = \frac{\partial \phi}{\partial q}, \qquad \mu = \frac{\partial \phi}{\partial \theta}. \tag{3.5}$$

Hamilton's equations now become

$$\dot{\rho} = \frac{\partial \tilde{K}}{\partial q}, \qquad \dot{\mu} = \frac{\partial \tilde{K}}{\partial \theta}, \qquad \dot{q} = -\frac{\partial \tilde{K}}{\partial \rho}, \qquad \dot{\theta} = -\frac{\partial \tilde{K}}{\partial \mu}, \tag{3.6}$$

where the dot refers to differentiation with respect to the running *time* variable along the orbit.

Since $\tilde{K}$ is not a function of θ we immediately have

$$\mu = \hat{m}$$

where $\hat{m}$ is constant. We may therefore write

$$\tilde{K} = \tilde{K}(q, p); \qquad p^2 = \rho^2 + \hat{m}^2/q^2. \tag{3.7}$$

Unlike a true dynamical problem the *time* dependence is of no direct physical interest. Integration of the system (3.6) is completed by recognizing that $\rho(q)$ is implicitly defined by

$$\tilde{K}(q, p) = \lambda. \tag{3.8}$$

The phase equation becomes

$$\dot{\phi} = \rho(q)\dot{q} + \hat{m}\dot{\theta}, \tag{3.9}$$

or if we write

$$R(q) = \int \rho(q)\, dq, \tag{3.10}$$

then

$$\phi = R(q) + \hat{m}\theta. \tag{3.11}$$

In view of the way in which the phase enters (2.2),

$$\hat{m} = \varepsilon m, \tag{3.12}$$

where m is an integer. In mechanical terms, (3.12) states that the *angular momentum* is negligible unless the circular wavenumber, m, is relatively large.

Passing next to the determination of the amplitude, A, equation (2.13) becomes

$$\frac{1}{q}\frac{\partial}{\partial q}\left(qA^2\frac{\partial \tilde{K}}{\partial \rho}\right) = 0 \tag{3.13}$$

or

$$A \propto 1/\left(q\tilde{K}_\rho\right)^{1/2}. \tag{3.14}$$

Thus, to the present order, we have determined

$$\psi \sim \left\{q\tilde{K}_\rho\left(q, [\rho^2 + m^2\varepsilon^2/q^2]^{1/2}\right)\right\}^{-1/2}\exp[im\theta + iR(q)/\varepsilon]. \tag{3.15}$$

The above procedure breaks down when the denominator of the amplitude vanishes. Thus $q = 0$ is one such critical point. Other critical points are determined by $K_\rho = 0$. For q bounded away from zero the second term under the radical is negligible (unless m is large) and in such cases the critical points are determined by

$$\tilde{K}_\rho(q, \rho) = 0, \tag{3.16}$$

and hence by the turning points of the equivalent one-dimensional problem. Additional possible turning points can occur near the origin. To see this write

$$\tilde{K}_\rho\left(q, \left[\rho^2 + m^2\varepsilon^2/q^2\right]^{1/2}\right) = \tilde{K}_p(q, p)\frac{\rho}{p}.$$ (3.17)

Near the origin,

$$\tilde{K}(0, p) = \lambda$$ (3.18)

has the solution

$$p^2 = \left[\rho^2 + m^2\varepsilon^2/q^2\right] \approx p_0^2(\lambda)$$ (3.19)

where p_0 represents the intercept at $q = 0$. Therefore, ρ, and hence $\tilde{K}_\rho$, vanishes for $q = O(\varepsilon)$ unless εm is not negligible. In the latter case the inner turning points are $O(1)$.

Each critical point indicates the need for a further analysis. As we will see the two critical points in the neighborhood of the origin can be considered simultaneously unless $\varepsilon m = O(1)$. Before these repairs are made some specific cases are considered in the next section.

4. Visual models. A model kernel which arises in visual experiments is [21]

$$K = \frac{A}{\pi(1 + q^2/a)}\left\{\exp\left[\frac{-u^2}{1 + q^2}\right] - B\exp\left[\frac{-u^2}{b^2(1 + q^2)}\right]\right\},$$ (4.1)

where A, B, a, b are positive constants. This kernel is in the form of (1.9). In particular, we observe that for this model, as we move well away from the central region, the resolution varies linearly with distance and sensitivity varies with inverse distance squared. Both properties are supported by observation [22, 23, 24].

The Wigner transform of (4.1) is given by

$$\tilde{K} = A\left(\frac{1 + q^2}{1 + q^2/a}\right)\left\{\exp\left[-p^2(1 + q^2)/4\right] - Bb^2\exp\left[-b^2(1 + q^2)p^2/4\right]\right\}.$$ (4.2)

For purposes of illustration and discussion we first consider the excitatory part of (4.1),

$$k = \frac{1}{\pi(1 + q^2/a)}\exp\left[\frac{-u^2}{1 + q^2}\right],$$ (4.3)

and its Wigner transform,

$$\tilde{k} = \left(\frac{1 + q^2}{1 + q^2/a}\right)\exp\left[-p^2(1 + q^2)/4\right].$$ (4.4)

The form of $\tilde{k}$ implies that the spectrum is restricted to the unit interval (as discussed below $a < 1$)

$$0 \leqslant \lambda \leqslant 1.$$ (4.5)

The contour lines of

$$\tilde{k}(q, p) = \lambda$$ (4.6)

are given explicitly by

$$p^2 = -[4/(1 + q^2)]\ln(\lambda[(1 + q^2/a)/(1 + q^2)]). \tag{4.7}$$

A sketch of such contours is shown in Fig. 1. In plotting these we have found it convenient to allow the magnitude p to formally take negative as well as positive values. The figure depicts a case for which $a < 1$, which is the case in experiment [21]. (Otherwise no closed contours exist.) Referring to the figure, closed contours occur for

$$a < \lambda < 1, \tag{4.8}$$

and open contours for

$$0 \leqslant \lambda < a. \tag{4.9}$$

As will be seen, the former correspond to discrete eigenvalues and the latter clearly correspond to the continuous spectrum. Recall that p and q each represent magnitudes of two-vectors and hence Fig. 1 is a two-space projection of four-dimensional phase space.

Contour lines in the ρ-q plane are obtained from (4.7) on substitution of (3.19),

$$\rho^2 = -(m^2\varepsilon^2/q^2) - [4/(1 + q^2)]\ln(\lambda[(1 + q^2/a)/(1 + q^2)]). \tag{4.10}$$

The corresponding contours are sketched in Fig. 2. The contour for $m = 0$ is one of those shown in Fig. 1. It plays a special role and we denote it by

$$\rho(q, m = 0) = \rho_0(q) = -[4/(1 + q^2)]\ln(\lambda[(1 + q^2/a)/(1 + q^2)]). \tag{4.11}$$

Except for $q = O(\varepsilon)$ the contours (4.10) lie close to $\rho_0(q)$, unless $m \gg 1$ in which case the inner turning point is removed from the neighborhood of the origin as discussed above.
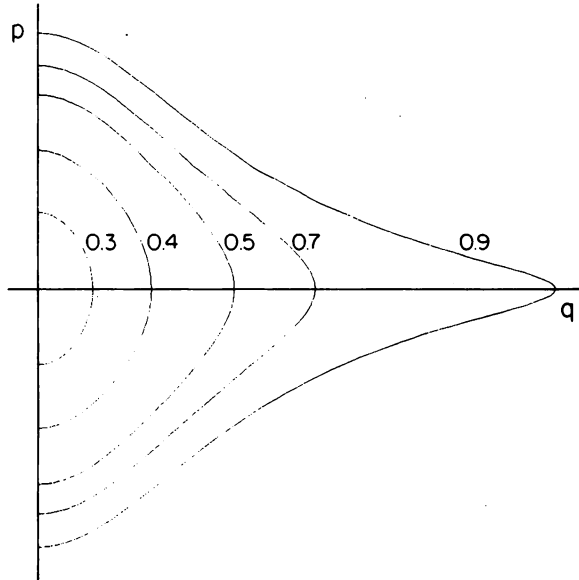

FIG. 1. Contours $p = p(q)$ for Eq. (4.7). Nominal values of $a = .25$ and $\varepsilon = .01$ are taken. The values of λ are indicated on the figure.
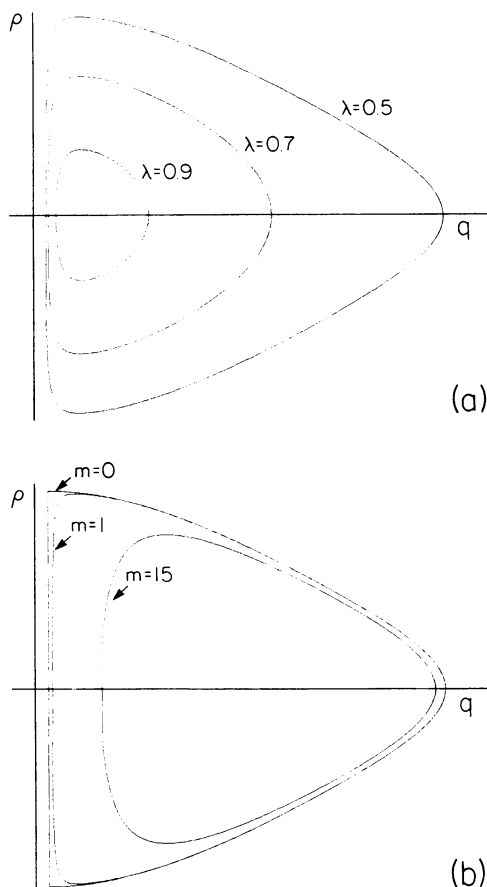
FIG. 2. Contours $\rho = \rho(q)$ for Eq. (4.10) with the nominal values of $a = .25$ and $\varepsilon = .01$. (a) indicates variation with λ, $m = 2$ fixed, (b) variation with m, $\lambda = .5$ fixed.

We return to the discussion of visual models and in particular to (4.2). Although an explicit form of $\rho(q; m)$ such as (4.10) does not seem feasible in this instance, the contours are easily obtained. Observe that q^2 as a function of $(1 + q^2)p^2$ is immediate and from this the required contours are easily generated. In Fig. 3 we give the result of such a procedure. Again we only focus on the case of closed contours and, as we show shortly, discrete eigenvalues. The figure caption contains a full discussion. We only remark on the fact that four branches are now present and therefore four terms will be present in the eigenfunction representation.

5. Turning points and other connections. In this section we consider the corrections which must be made at the critical points mentioned in the previous section. This will lead not only to a uniform description of the eigenfunctions but also to the determination of the eigenvalues. In approaching this portion of the problem we recognize that the nature of the contours of $\tilde{K}$ will play an essential role. It is impractical to anticipate all possible

topologies. We will therefore be content to consider two examples which incorporate enough features to deal with fairly complicated topologies. These will permit us to treat the case depicted in Fig. 2 which we term type I and that depicted in Fig. 3 which we term type II.

Type I. In a type I situation we have two branches for all values of m, as indicated in Fig. 2. For the turning point at the right, q_0 of the figure, $\rho = 0$. As is well known [1, 16, 17], the phase advances by $\pi/2$ as we pass clockwise around a turning point. Thus we may
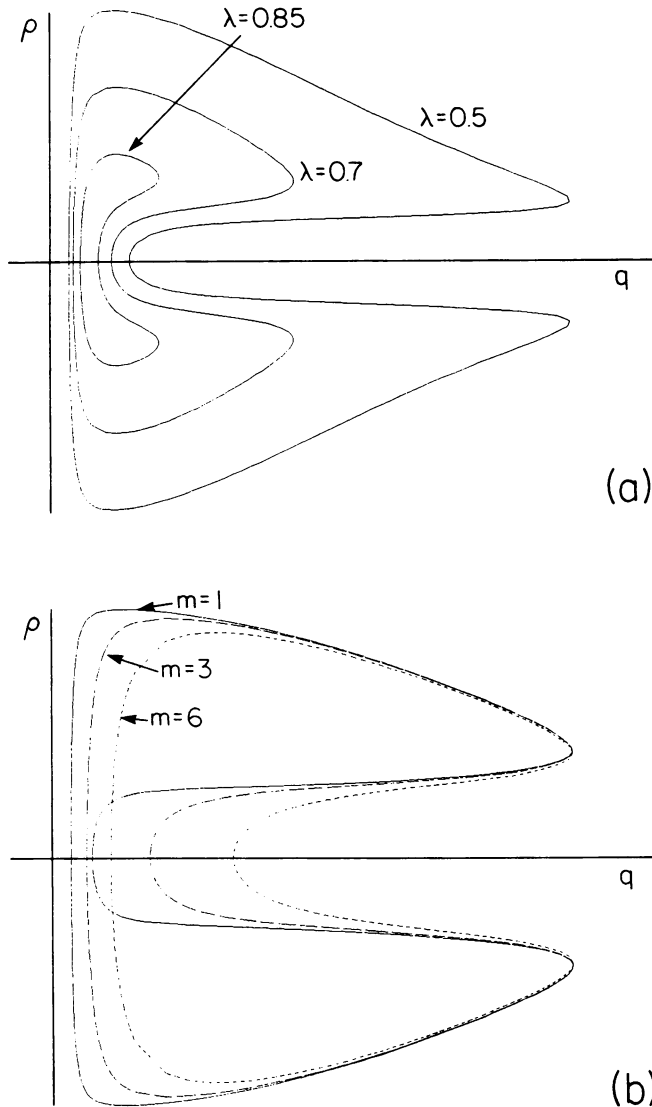

FIG. 3. Contours illustrate Eq. (4.2). The values $a = .395$, $\varepsilon = .009$, $b^2 = 49$ represent values based on human psychophysical data. The nominal value of $m = 4$ is taken in (a) to illustrate variation with λ, while $\lambda = .7$ in (b) with m varying.

represent the WKB form of the solution by

$$\psi_w = \exp(im\theta)\left[\frac{\tilde{K}_p^0}{p_0 q \pi \tilde{K}_\rho}\right]^{1/2} \sin\left(\frac{1}{\varepsilon}\int_q^{q_0}\rho(s)\,ds + \frac{\pi}{4}\right). \tag{5.1}$$

p_0 is the intercept of $\rho_0(p)$ as indicated in the figure and

$$\tilde{K}_p^0 = \tilde{K}_p(0, p_0).$$

Multiplicative constants have been introduced for later convenience. Eq. (5.1) has been written so as to accommodate the phase condition at q_0. If the inner turning point Q_0 is bounded away from the origin (so that $m \gg 1$) and indicated by the curve with the hatched interior, then the phase condition at Q_0 yields

$$A(\lambda) = 2\int_{Q_0}^q \rho(s)\,ds = (2n+1)\pi\varepsilon. \tag{5.2}$$

Thus $A(\lambda)$ is the hatched area of Fig. 2 and n is an integer. As indicated, A is a function of λ and in fact (5.2) is the condition that λ be an eigenvalue. We have termed this condition the area rule [1, 2, 3]. The neighborhood of the origin proves to be of no consequence and we comment on this below.

If m does not become large the inner turning point is in the neighborhood of the origin, and from (3.19),

$$Q_0 = O(\varepsilon). \tag{5.3}$$

Inner analysis. To deal with the turning point (5.3) as well as the critical point at the origin, $q = 0$, we consider the inner limit: q/ε^μ fixed; $0 < \mu < 1$; $\varepsilon \downarrow 0$. Observe that under this limit $\rho \to p_0$, $\tilde{K}_\rho \to \tilde{K}_\rho^0$. We will also need the limit of

$$\int_q^{q_0}\rho(s)\,ds = \int_{Q_0}^{q_0}\rho(s)\,ds - \int_{Q_0}^q\rho(s) = \frac{A}{2} - \int_{Q_0}^q\rho(s)\,ds. \tag{5.4}$$

As before A refers to the area enclosed by the curve described by (4.10) (the dashed curve in Fig. 4) and is therefore still a function of ε. A straightforward calculation shows that the cross-hatched area of Fig. 4 is $m\pi\varepsilon$—independently of p_0! It therefore follows that

$$A \sim A_0 - m\pi\varepsilon \quad \text{and} \quad \int_{Q_0}^q\rho(s)\,ds \sim p_0 q - m\pi\varepsilon, \tag{5.5}$$

where A_0 is the area of the entire hatched portion of Fig. 4 or, equivalently, it is the area enclosed by ρ_0, (4.11). It therefore follows that the inner limit of (5.1) is given by

$$\psi_w \sim \frac{\exp(im\theta)}{(\pi q p_0)^{1/2}}\sin\left(\frac{A_0}{2\varepsilon} - p_0 q + \frac{\pi}{4}\right). \tag{5.6}$$

Under the inner limit the integral equation itself (2.1) becomes

$$\int K(0, \mathbf{x} - \mathbf{y})\psi_I(\mathbf{y})\,d\mathbf{y} = \lambda\psi_I(\mathbf{x}). \tag{5.7}$$

Thus

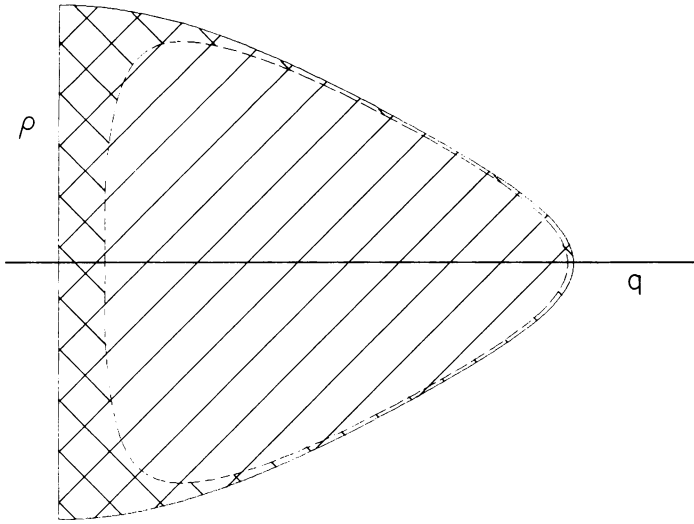$$\psi_I \propto \exp(-i\mathbf{p}\cdot\mathbf{x}) \tag{5.8}$$

FIG. 4. Nominal values of $a = .25$, $\varepsilon = .05$ and $\lambda = .5$ have been used in this sketch. The continuous line corresponds to $m = 0$ and the dashed line to $m = 2$.

is a solution if **p** is such that

$$\tilde{K}(0, p) = \lambda \tag{5.9}$$

and hence

$$p = p_0. \tag{5.10}$$

Superposition of all admissible solutions gives the inner solution

$$\psi_I = \frac{1}{2\pi} \int_0^{2\pi} f(\theta') \exp\left[-ip_0 r \cos(\theta - \theta')\right] d\theta'$$

$$= \frac{1}{2\pi} \int_0^{2\pi} f(\theta') \exp\left[-i\frac{p_0 q}{\varepsilon} \cos(\theta - \theta')\right] d\theta', \tag{5.11}$$

where $f(\theta)$ is an arbitrary function. Under the outer limit (q fixed and $\varepsilon \downarrow 0$) a simple stationary point calculation yields

$$\psi_I \sim \left[\varepsilon/(p_0 q 2\pi)\right]^{1/2} \left\{ f(\theta)\exp\left(-\frac{ip_0 q}{\varepsilon} + \frac{i\pi}{4}\right) + f(\theta - \pi)\exp\left(\frac{ip_0 q}{\varepsilon} - \frac{i\pi}{4}\right)\right\}. \tag{5.12}$$

Comparison of (5.12) with (5.6) implies that $f \propto \exp(im\theta)$ and for convenience we write

$$f(\theta) = \frac{C}{i(2\varepsilon)^{1/2}} \exp(im\theta). \tag{5.13}$$

Therefore, in order to obtain a match between the inner and outer solutions, we must have

$$C = \exp\left[iA_0/(2\varepsilon)\right] \tag{5.14}$$

and

$$C\exp(-im\pi) = -\exp(-iA_0/(2\varepsilon)). \tag{5.15}$$

It therefore follows that

$$A_0 = (2n + 1 - m)\pi\varepsilon \tag{5.16}$$

where n is an integer. If (5.5) is substituted we obtain

$$A(\lambda) = (2n + 1)\pi\varepsilon \tag{5.17}$$

where n is an integer so that the area rule is again obtained.

Next the constant C is given by

$$C = \exp\left[i\left(n - \frac{m}{2} + \frac{1}{2}\right)\pi\right]. \tag{5.18}$$

If (5.13) is substituted into (5.11) we see that the inner form of the eigenfunction can be evaluated in terms of a Bessel function and is

$$\psi_{\mathrm{I}} = \frac{(-)^{n+m} J_m(p_0 q/\varepsilon)}{\sqrt{2\varepsilon}} \exp(im\theta). \tag{5.19}$$

We mention in passing that the eigenfunction in the neighborhood of the origin must also be of this form even when $m\varepsilon \geqslant O(1)$. Since this implies that $m \gg 1$, it follows from the properties of the Bessel function that ψ is negligible to the left of the inner turning point in this instance.
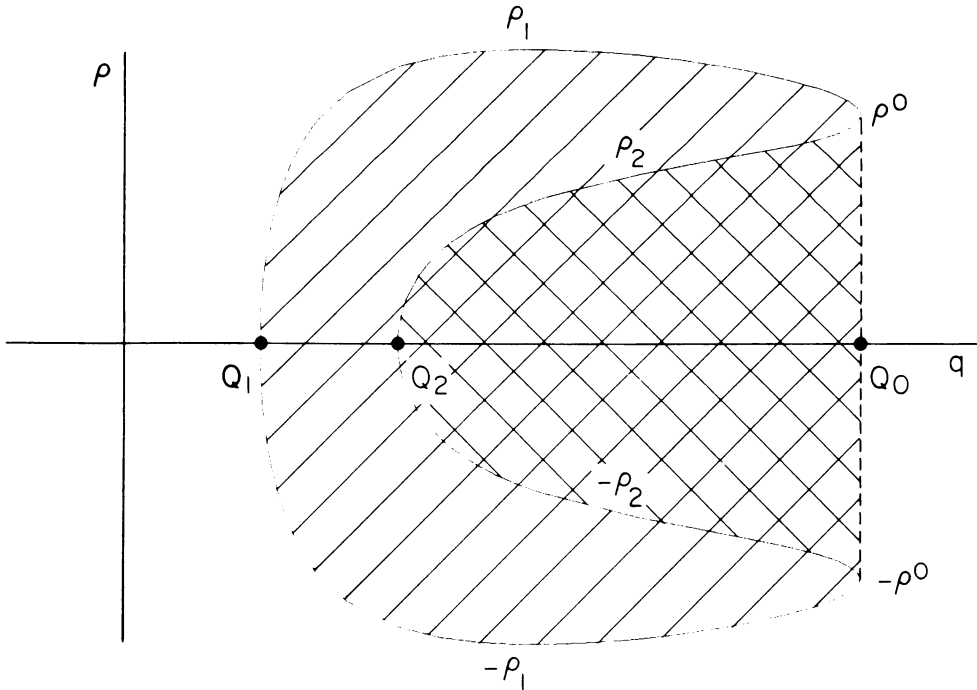

FIG. 5. Curve is based on Eq. (4.2) with $\varepsilon = .009$, $a = .395$, $\lambda = .85$, and $m = 4$. See text for discussion.

Type II. The type II situation is characterized by the curves of Fig. 3. For later purposes we show two additional figures, Fig. 5 and Fig. 6, which depict this case. In both cases the four branches are denoted by $\pm\rho_1$, $\pm\rho_2$, the outer turning points by $(q_0, \pm\rho^0)$, and the two inner turning points by Q_1 and Q_2. Fig. 5 describes the case of the inner turning points bounded away from the origin, $m\varepsilon \gg 1$, while for Fig. 6, $m\varepsilon \ll 1$, so the Q_1 and Q_2 lie in the neighborhood of the origin.

The WKB or outer form of the eigenfunction is given by

$$
\psi_w = \frac{\exp(im\theta)}{2\sqrt{q\pi}} \left[a \left\{ \frac{\exp\left[\frac{i}{\varepsilon}\int_q^{q_0}\rho_1(s)\,ds + i\pi/4\right]}{\left(\tilde{K}_\rho(q,\rho_1)\right)^{1/2}} - \frac{\exp\left[\frac{i}{\varepsilon}\int_q^{q_0}\rho_2(s)\,ds - i\pi/4\right]}{\left(\tilde{K}_\rho(q,\rho_2)\right)^{1/2}} \right\} \right.
$$
$$
\left. + b \left\{ \frac{\exp\left[-\frac{i}{\varepsilon}\int_q^{q_0}\rho_2(s)\,ds + i\pi/4\right]}{\left(\tilde{K}_\rho(q,\rho_2)\right)^{1/2}} - \frac{\exp\left[-\frac{i}{\varepsilon}\int_q^{q_0}\rho_1(s)\,ds - i\pi/4\right]}{\left(\tilde{K}_\rho(q,\rho_1)\right)^{1/2}} \right\} \right], \quad (5.20)
$$

where the constants a and b remain to be determined. In writing (5.20) we have already taken care of the requirement that the phase advance by $\pi/2$ in clockwise passage around each of the two outer turning points.

We first treat the case depicted in Fig. 5 ($m\varepsilon = O(1)$). Since the inner turning points are well away from the origin in this case the standard treatment applies. The condition that the phase advance by $\pi/2$ in advancing across Q_1 and across Q_2 in the clockwise direction
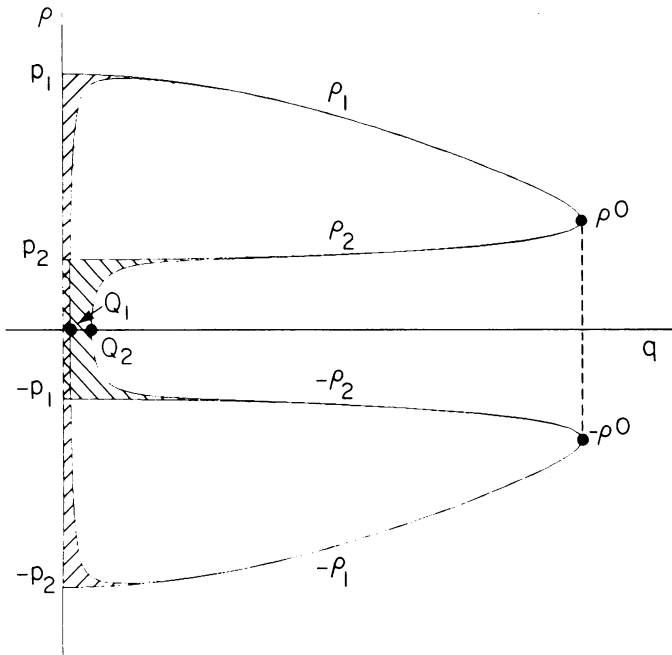

FIG. 6. Curve is based on Eq. (4.2) with $\varepsilon = .009$, $a = .395$, $\lambda = .7$, $m = 3$. See text.

yields

$$A(\lambda) = 2\left(\int_{Q_1}^{q_0} \rho_1(s)\, ds - \int_{Q_2}^{q_0} \rho_2(s)\, ds \right) = (A_1 - A_2) = (2n+1)\pi\varepsilon, \quad (5.21)$$

where A is the area enclosed by the contour of Fig. 5, i.e., it is the entire hatched area A_1 minus the cross-hatched area A_2. As could have been anticipated, the area rule for the determination of λ, (5.21), is still valid. We also find

$$\frac{a}{b} = \frac{1}{2}\left[\exp(-iA_2/\varepsilon) - \exp(-iA_1/\varepsilon) \right], \quad (5.22)$$

which finishes the determination of the WKB form of the eigenfunction.

In order to treat the case depicted in Fig. 6 an inner analysis is required.

Inner analysis. The same scaling used in the type I situation applies again. Under the inner limit we find

$$\int_q^{q_0} \rho_j(s)\, ds \sim \frac{A_j^0}{2} - p_j q = \frac{A_j + m\pi\varepsilon}{2} - p_j q; \quad j = 1, 2, \quad (5.23)$$

where the arguments follow those given for (5.4) and (5.5). A_1^0 and A_2^0 refer to the areas obtained when $\varepsilon = 0$ (or $m = 0$). As is seen in Fig. 6,

$$A_0 = A_1^0 - A_2^0 \quad (5.24)$$

refers to the area of the two unconnected lobes. It then follows that (5.20) under the inner limit becomes

$$\psi_w \sim \frac{\exp(im\theta)}{2\sqrt{q\pi}}\left[a\left\{ \frac{\exp\left[(iA_1^0/(2\varepsilon)) - ip_1 q + i\pi/4\right]}{\left(\tilde{K}_\rho(p_1,0)\right)^{1/2}} - \frac{\exp\left[(iA_2^0/(2\varepsilon)) - ip_2 q - i\pi/4\right]}{\left(\tilde{K}_\rho(p_2,0)\right)^{1/2}} \right\}\right.$$

$$\left. + b\left\{ \frac{\exp\left[-(iA_2^0/(2\varepsilon)) + ip_2 q + i\pi/4\right]}{\left(\tilde{K}_\rho(p_2,0)\right)^{1/2}} - \frac{\exp\left[-(iA_1^0/(2\varepsilon)) + ip_1 q - i\pi/4\right]}{\left(\tilde{K}_\rho(p_1,0)\right)^{1/2}} \right\}\right]. \quad (5.25)$$

The inner form of the integral equation is again given by (5.7) and its solution has the form (5.8), with p such that (5.9) is satisfied. However, instead of (5.10) we now find

$$p = p_1, p_2. \quad (5.26)$$

Thus superposition yields

$$\psi_I = \frac{1}{2\pi}\int_0^{2\pi} f_1(\theta') \exp\left[-\frac{ip_1 q}{\varepsilon}\cos(\theta' - \theta) \right] d\theta'$$

$$+ \frac{1}{2\pi}\int_0^{2\pi} f_2(\theta')\exp\left[-\frac{ip_2 q}{\varepsilon}\cos(\theta' - \theta) \right] d\theta' \quad (5.27)$$

where the functions f_1, f_2 remain to be determined from an outer matching. Under the outer limit (5.27) becomes

$$\psi_I \sim \left[\varepsilon/(p_1 q 2\pi) \right]^{1/2}\left\{ f_1(\theta)\exp\left[-\frac{ip_1 q}{\varepsilon} + \frac{i\pi}{4} \right] + f_1(\theta - \pi)\exp\left[\frac{ip_1 q}{\varepsilon} - \frac{i\pi}{4} \right] \right\}$$

$$+ \left[\varepsilon/(p_2 q 2\pi) \right]^{1/2}\left\{ f_2(\theta)\exp\left[-\frac{ip_2 q}{\varepsilon} + \frac{i\pi}{4} \right] + f_2(\theta - \pi)\exp\left[\frac{ip_2 q}{\varepsilon} - \frac{i\pi}{4} \right] \right\}. \quad (5.28)$$

Comparison of (5.28) with (5.25) implies

$$f_1(\theta) = C_1 e^{im\theta} (1/(2\varepsilon))^{1/2},$$
$$f_2(\theta) = C_2 e^{im\theta} (1/(2\varepsilon))^{1/2}. \tag{5.29}$$

It then follows that

$$A_0 = (2n + 1)\varepsilon\pi = A(\lambda) = A_1 - A_2. \tag{5.30}$$

Once A is fixed by (5.30), A_1 and A_2 are each determined as a consequence of their functional relationship as required by $\tilde{K}(\rho, q) = \lambda$. The ratio of a to b is again determined by (5.22) and

$$\begin{cases} C_1 = a\left(p_1/\tilde{K}_\rho^0(p_1) \right)^{1/2} \exp\left[\dfrac{iA_1^0}{2\varepsilon} \right], \\[2mm] C_2 = (-)^m a\left(p_2/\tilde{K}_\rho^0(p_2) \right)^{1/2} \exp\left[\dfrac{iA_1^0}{2\varepsilon} \right] \end{cases} \tag{5.31}$$

determine the remaining constants.

To finish the job, (5.29) is substituted into (5.27), which again results in Bessel functions of order m,

$$\psi_I = \frac{\exp(im\theta)}{(2\varepsilon)^{1/2}} \left[a\left(\frac{p_1}{\tilde{K}_\rho^0(p_1)} \right)^{1/2} J_m(p_1 q/\varepsilon)\exp\left(\frac{iA_1}{2\varepsilon} \right) \right.$$
$$\left. + ib\left(\frac{p_2}{\tilde{K}_\rho^0(p_2)} \right)^{1/2} J_m(p_2 q/\varepsilon)\exp\left(\frac{-iA_2}{2\varepsilon} \right) \right]. \tag{5.32}$$

6. Principal eigenfunctions. For eigenfunctions of low index, the so-called principal eigenfunctions, last section's approach is no longer reliable. To see this observe that the areas calculated say in (5.2) and (5.14) are $O(\varepsilon)$ and as a result, the phase term no longer dominates the amplitude term, so the basis of the expansion procedure becomes questionable. This problem can be treated by going directly to the integral equation and introducing appropriate approximations. [See Section 4 of ref. 1 for this approach in the one-dimensional case.] In this section we approach the problem of principal eigenfunctions in a somewhat different, but equivalent fashion.

It will be useful in the sequel to start with the following example:

$$-\nabla^2\phi + \varepsilon^2 r^2\phi = -\nabla^2\phi + \varepsilon^2(x^2 + y^2)\phi = \lambda\phi. \tag{6.1}$$

(As will be clear in a moment, ε scales out of the problem). This is a separable equation and is easily seen to have the solution

$$\phi = H_k(\varepsilon^{1/2}y)H_n(\varepsilon^{1/2}x)\exp\left[-\frac{\varepsilon}{2}(x^2 + y^2) \right]; \qquad \lambda = \{2n + 2k + 2\}\varepsilon, \tag{6.2}$$

where H_n denotes the nth Hermite polynomial [25]. The eigenvalues have the degeneracy associated with the integer $(n + k)$ being constant. Another form of the eigentheory is

obtained by introducing cylindrical coordinates. Thus we write

$$\phi = \ell(r)\exp[im\theta], \tag{6.3}$$

in which case (6.1) becomes

$$-\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{m^2}{r^2} - \varepsilon^2 r^2\right)\ell = \lambda\ell \tag{6.4}$$

and

$$\ell = r^m L_n^{(m)}(\varepsilon r^2)\exp[-\varepsilon r^2/2], \qquad \lambda = (4n + 2m + 2)\varepsilon, \tag{6.5}$$

where $L_n^{(m)}$ represents the modified Laguerre polynomial [25].

The Wigner transform of the operator associated with (6.1) is $\tilde{K} = p^2 + q^2$ so that the contours are simply the circles

$$p^2 + q^2 = \lambda \tag{6.6}$$

or more precisely semicircles for $q > 0$. This corresponds to a type I contour and, if we apply the rule (5.11), we see that in this instance it exactly gives the eigenvalues (6.5). For low index however, the approximate eigenfunction does not give a good fit to the exact form (6.5), a fact already foreshadowed by the one-dimensional treatment [1, 2, 3].

In the remainder of this section we treat the two cases referred to as types I and II in the previous section. We now adopt a slight change of notation and write the kernels as

$$\tilde{K} = \tilde{K}(p^2, q^2), \tag{6.7}$$

a form which underlines the symmetries assumed thus far. The corresponding contours are given by

$$\tilde{K}(p^2, q^2) = \lambda. \tag{6.8}$$

Type I *principal eigenfunctions.* Under investigation are those cases for which the contours of (6.8) enclose small areas. Of necessity these lie near the peaks of (6.8). In the present instance the only peak lies at the origin. In this neighborhood we can Taylor expand to obtain

$$\tilde{K} \sim \lambda_0 - k_1 p^2 - k_2 q^2 \tag{6.9}$$

where

$$\lambda_0 = \tilde{K}(0,0), \qquad k_1 = -\tilde{K}_{,1}(0,0), \qquad k_2 = -\tilde{K}_{,2}(0,0). \tag{6.10}$$

The negative signs in (6.10) reflect the fact that the origin is a peak, i.e., $k_1, k_2 > 0$. If we use (6.9) in (6.8) then

$$\lambda_0 - \lambda = \lambda' = k_1 p^2 + k_2 q^2. \tag{6.11}$$

Equation (6.11) can be regarded in its own right as arising out of the eigenfunction problem

$$-k_1\nabla^2\phi + k_2\varepsilon^2 r^2\phi = \lambda'\phi. \tag{6.12}$$

The Wigner transform of the operator in (6.12) is (6.11) and it shares with the original integral equation (2.1) for the particular kernel implicit in (6.12) the same contour structure in the neighborhood of the origin. Intuition would then say that (6.12) has a

principal eigenstructure which closely approximates that of the original problem. This can be confirmed by directly investigating the integral equation for principal eigenfunctions—an approach taken in the one-dimensional case and therefore not deemed necessary here [1, 2].

Equation (6.12) is easily scaled into the form (6.1) and the result yields

$$\phi = r^m L_n^{(m)}\left(\frac{k_2 \epsilon r^2}{k_1}\right) \exp\left[im\theta - \frac{\epsilon k_2 r^2}{2k_1}\right],\tag{6.13}$$

$$\lambda = \lambda_0 - k_1(4n + 2m + 2)\epsilon.\tag{6.14}$$

In particular, for the simplified kernel (4.3) we have

$$1 - \lambda = \frac{1}{4}p^2 + \left(\frac{1-a}{a}\right)q^2\tag{6.15}$$

with the resulting eigenfunctions and eigenvalues,

$$\phi = r^m L_n^{(m)}\left(4\left(\frac{1-a}{a}\right)\epsilon r^2\right)\exp\left[im\theta - 2\left(\frac{1-a}{a}\right)\epsilon r^2\right],\tag{6.16}$$

$$\lambda = \left[1 - (1/4)(4n + 2m + 2)\epsilon\right].\tag{6.17}$$

Type II *principal eigenfunctions.* In this case the Wigner transform has two peaks located at

$$(p,q) = (\pm p_0, 0)\tag{6.18}$$

with p_0 such that

$$\tilde{K}_{,1}(p_0^2, 0) = 0.$$

The appropriate Taylor expansion of (6.8) has the form

$$\tilde{K} \sim \lambda_0 - \alpha(p^2 - p_0^2)^2 - \beta q^2\tag{6.19}$$

with

$$\alpha = -\tilde{K}_{,1}(p_0^2, 0), \qquad \beta = -\tilde{K}_{,2}(p_0^2, 0).$$

In particular, for the visual model (4.2) we have

$$\tilde{K} \sim A\exp\left[-p_0^2/4\right](1 - 1/b^2)\left\{1 - (1-a)q^2/a - b^2(p^2 - p_0^2)^2/32\right\}\tag{6.20}$$

with p_0 such that

$$p_0^2 = 4\ln\frac{Bb^4}{b^2 - 1}.$$

If (6.19) is substituted into (6.8) we obtain

$$\frac{\lambda_0 - \lambda}{\alpha} = (p^2 - p_0^2)^2 + (\beta/\alpha)q^2.\tag{6.21}$$

Considered in its own right, (6.21) is the Wigner transform of the operator for the following eigenproblem:

$$\delta\Lambda\phi = (\nabla^2 + p_0^2)^2\phi + \delta^2 r^2\phi,\tag{6.22}$$

where

$$\delta^2 = \frac{\beta}{\alpha}\varepsilon^2, \qquad \Lambda = \frac{\lambda_0 - \lambda}{\alpha\delta}. \tag{6.23}$$

The factor δ on the left hand, which has been introduced for formal reasons, is appropriate as indicated by the previous case. Unlike the previous case (type I) this principal eigenproblem does not lead to a solvable canonical equation and instead we must resort to asymptotics in the limit $\delta \downarrow 0$.

The gross features of the appropriate approach follow from a study of the comparable one-dimensional problem and we digress to consider

$$\delta\Lambda\phi = \left(\frac{d^2}{dx^2} + p_0^2\right)^2\phi + \delta^2 x^2\phi. \tag{6.24}$$

A naive perturbation analysis applied to (6.24) soon suggests the two-scale procedure

$$\phi = \phi(x_0, x_1; \delta), \qquad x_0 = x, \qquad x_1 = \delta^{1/2}x \tag{6.25}$$

so that (6.24) becomes

$$\delta\Lambda\phi = \left[\left(\frac{\partial}{\partial x_0} + \delta^{1/2}\frac{\partial}{\partial x_1}\right)^2 + p_0^2\right]^2\phi + \delta x_1^2\phi. \tag{6.26}$$

This in turn suggests the expansion

$$\phi = \sum_{n=0}^{\infty} \delta^{n/2}\phi_n(x_0, x_1). \tag{6.27}$$

At the lowest order,

$$\left(\frac{\partial^2}{\partial x_0^2} + p_0^2\right)^2\phi_0 = 0 \tag{6.28}$$

with

$$\phi_0 = f(x_1)\exp[ip_0 x_0]. \tag{6.29}$$

The conjugate need not be considered and the remaining solutions may be excluded since they show growth at infinity. At the next order, we find $\phi_1 = 0$ and at $O(\delta)$ we are left with

$$\left(\frac{\partial^2}{\partial x_0^2} + p_0^2\right)^2\phi_2 = \left(\Lambda + 4p_0^2\frac{\partial^2}{\partial x_1^2} - x_1^2\right)(f(x_1)\exp[ip_0 x_0]). \tag{6.30}$$

Secular growth is avoided by choosing f such that

$$-4p_0^2\frac{\partial^2}{\partial x_1^2}f + x_1^2 f = \Lambda f. \tag{6.31}$$

The solution to (6.31) is

$$\Lambda = 2p_0(2n + 1), \tag{6.32}$$

$$f = H_n\left(\frac{x_1}{(2p_0)^{1/2}}\right)\exp\left[-\frac{x_1^2}{4p_0}\right] = H_n\left(x\left(\frac{\delta}{2p_0}\right)^{1/2}\right)\exp\left(-\frac{x^2\delta}{4p_0}\right), \tag{6.33}$$

and the asymptotic form of the eigenfunction is obtained by substituting (6.33) into (6.29).

The corresponding analysis of (6.22) parallels the above one-dimensional treatment, and we simply outline the steps. We first separate variables

$$\phi = \Phi \exp[im\theta] \tag{6.34}$$

and also write

$$\nabla^2 = \frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{m^2}{r^2}. \tag{6.35}$$

A straightforward perturbation procedure again indicates the need for a two-scale analysis and in anticipation of this we define

$$r_0 = r, \qquad r_1 = \delta^{1/2} r \tag{6.36}$$

and hence

$$\frac{d}{dr} = \frac{\partial}{\partial r_0} + \delta^{1/2}\frac{\partial}{\partial r_1}. \tag{6.37}$$

The presence of purely geometrical terms in (6.35) and (6.22) leads to some flexibility in the expansion of the operators. A convenient form of expansion is given by

$$\left(\nabla^2 + p_0^2\right)^2 + \delta^2 r^2 = \left(D_0^2\right)^2 + \delta^{1/2}4\frac{\partial}{\partial r_1}\frac{\partial}{\partial r_0}D_0^2$$

$$+ \delta\left\{ r_1^2 + 4p_0^2\frac{\partial}{\partial r_1^2} + \left(6\frac{\partial}{\partial r_1^2} + \frac{2}{r_1}\frac{\partial}{\partial r_1}\right)D_0^2\right\} + O(\delta^{3/2}) \tag{6.38}$$

where

$$D_0^2 = \frac{\partial}{\partial r_0^2} + \frac{1}{r_0}\frac{\partial}{\partial r_0} - \frac{m^2}{r_0^2} + p_0^2. \tag{6.39}$$

Then if we expand

$$\Phi = \Phi(r_0, r_1\delta) = \sum \delta^{n/2}\Phi_n(r_0, r_1), \tag{6.40}$$

insert it and (6.38) into (6.22) to lowest order, we obtain

$$\left(D_0^2\right)\Phi_0 = 0, \tag{6.41}$$

the solution of which is

$$\Phi_0 = J_m(p_0 r_0)f(r_1). \tag{6.42}$$

Other solutions of (6.41) have unacceptable growth properties.

At the next order we find that $\Phi_1 = 0$, and at $O(\delta)$ we obtain the condition that f be such that

$$\left(-4p_0^2\frac{d^2}{dr_1^2} + r_1^2\right)f = \Lambda f. \tag{6.43}$$

This is the same as Eq. (6.31) and has the solution (6.32), (6.33) with x_1 replaced by r_1. If we remove all the intermediate scale changes and reassemble the asymptotic solution the result is

$$\lambda \sim \lambda_0 - 2\varepsilon\sqrt{\alpha\beta}(2n + 1)p_0, \tag{6.44}$$

and

$$\phi = J_m(p_0 r) H_n\left(\left[\frac{\beta \varepsilon^2}{4\alpha p_0^2}\right]^{1/4} r\right) \exp\left[im\theta - \frac{1}{2}\left(\frac{\beta \varepsilon^2}{4\alpha p_0^2}\right)^{1/2} r^2\right]. \qquad (6.45)$$

The dependence of λ, (6.44), on the angular quantum number, m, does not appear until next order, in which case it can be shown that

$$\lambda \sim \lambda_0 = 2\varepsilon\sqrt{\alpha\beta}(2n+1)\beta_0 - \varepsilon\beta\frac{4m^2-1}{4p_0^2}. \qquad (6.46)$$

REFERENCES

[1] L. Sirovich and B. W. Knight, *On the eigentheory of operators which exhibit a slow variation*, Quart Appl. Math. **38**, 469–488 (1981)

[2] L. Sirovich and B. W. Knight, *Contributions to the eigenvalue problem for slowly varying operators*, SIAM J. Appl. Math. **42**, 356–377 (1982)

[3] B. W. Knight and L. Sirovich, *The Wigner transform and some exact properties of linear operators*, SIAM J. Appl. Math. **42**, 378–389 (1982)

[4] L. Sirovich and B. W. Knight, *The eigenfunction problem in higher dimensions*, I. *Asymptotic theory*, Proc. Natl. Acad. Sci. USA **82**, 8275–8278 (1985)

[5] B. W. Knight and L. Sirovich, *The eigenfunction problem in higher dimensions* II. *Exact Results*, Proc. Natl. Acad. Sci. USA **83**, 527–530 (1986)

[6] T. Cornsweet, *Visual Perception*, Academic Press, New York, 1970

[7] F. Ratliff, *Mach Bands: Qualitative Studies on Neural Networks in the Retina*, Holden-Day, San Francisco, 1965

[8] F. W. Campbell and J. G. Robson, *Application of Fourier analysis to the visibility of gratings*, J. Phys. (London) **197**, 551–566 (1968)

[9] N. Graham and J. Nachmias, *Detecting of grating patterns containing two spatial frequencies: A comparison of single and multiple channel models*, Vis. Res. **11**, 251–261 (1971)

[10] J. G. Robson and N. Graham, *Probability summation and regional variation in contrast sensitivity across the visual field*, Vis. Res. **21**, 409–418 (1981)

[11] H. R. Wilson and J. R. Bergen, *A four mechanism model for threshold spatial vision*, Vis. Res. **19**, 19–32 (1979)

[12] S. W. Kuffler, *Discharge patterns and functional organization of mammalian retina*, J. Neurophysiol. **16**, 37–68 (1953)

[13] R. W. Rodieck, *Qualitative analysis of cat retinal cell response to visual stimuli*, Vis. Res. **5**, 583–601 (1965)

[14] E. P. Wigner, *On the quantum correction for thermodynamic equilibrium*, Phys. Rev. **40**, 749–759 (1932)

[15] F. John, *Partial Differential Equations*, Springer-Verlag, New York, 1978

[16] J. Heading, *An Introduction to Phase Integral Methods*, Methuen & Co., Ltd., London, 1962

[17] C. M. Bender and S. Orzag, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1978

[18] J. B. Keller, *Corrected Bohr-Sommerfeld quantum conditions for non-separable systems*, Ann. Phys. **9**, 180–188 (1958)

[19] J. B. Keller, *Semiclassical Mechanics*, SIAM Review **27**, 485–504 (1985)

[20] H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, Massachusetts, 1981

[21] L. Sirovich and B. W. Knight, *Characteristic patterns of an inhomogeneous imaging system with applications to vision*, J. Op. Soc. Amer. **A3**, 358–364 (1986)

[22] Y. T. So and R. Shapley, *Spatial properties of X and Y cells in the lateral geniculate nucleus of the cat and conduction velocities of thin inputs*, Exp. Brain Res. **36**, 533–550 (1979)

[23] B. G. Cleland, T. H. Harding, and U. Tulunay-Keesey, *Visual resolution and receptive field size: examination of two kinds of retinal ganglion cells*, Science **205**, 1015–1017 (1979)

[24] B. Fischer and H. V. May, *Invarianzen in der katzen retina: gesetzmassig beziehungen zwischen empfindlichkeit, grosse und lage receptivier felder von ganglienzellen*, Exp. Brain Res. **11**, 448–464 (1970)

[25] W. Magnus, F. Oberhettinger and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer-Verlag, New York, 1966