HIGHER CORRECTIONS TO THE WKB APPROXIMATION

by N. G. VAN KAMPEN

Instituut voor Theoretische Fysica der Rijksuniversiteit te Utrecht, Nederland

Synopsis
    A systematic method of successive approximations is described, of which the first
step is the WKB approximation. The higher corrections consist of successive recombi-
nations of the two WKB waves, rather than a perturbation expansion in their in-
teraction. At the same time the local wave number (or refractive index) undergoes
successive corrections. Using this method it is shown that the quantummechanical
reflection by a potential barrier is a small quantity of order $\hbar^\infty$, provided that the
barrier is smooth, and so low that there is no classical turning point.

1. *Introduction.* The Schrödinger equation for a spinless particle in one
dimension is

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2} + \frac{2m}{\hbar^2}\{E - V(x)\}\,\varphi = 0.$$

We shall suppose throughout that $E > V(x)$ for all x, which means that
classical turning points are excluded. We may then use the abbreviation

$$p(x) = \sqrt{2mE - 2mV(x)} > 0;$$

$p(x)$ is the momentum that the particle should have at the point x according
to classical mechanics. It is convenient to define two components

$$u_1(x) = \varphi(x), \quad u_2(x) = \hbar\,\frac{\mathrm{d}\varphi(x)}{\mathrm{d}x},$$

and to write the Schrödinger equation in the matrix form[1])

$$\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \frac{1}{\hbar}\begin{pmatrix} 0 & 1 \\ -p^2(x) & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \frac{1}{\hbar}\,A(x)\,u.$$

For any fixed value of x the matrix A can be diagonalized, i.e., a non-
singular matrix $P(x)$ and a diagonal matrix $\Lambda(x)$ can be found such that

$$P(x)^{-1}\,A(x)\,P(x) = \Lambda(x), \qquad \Lambda \text{ diagonal.}$$

(Our condition $p^2 > 0$ guarantees that there are two different eigenvalues.)

The explicit matrices Λ and P are

$$\Lambda = \begin{pmatrix} ip & 0 \\ 0 & -ip \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 1 \\ ip & -ip \end{pmatrix}.$$

If the vector u is transformed accordingly

$$u = Pv,$$

the new vector v obeys the equation

$$\frac{dv}{dx} = \frac{ip(x)}{\hbar} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} v - \frac{p'(x)}{2p(x)} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} v. \tag{1}$$

The second term on the right is small compared to the first one, since it does not have an $\hbar$ in the denominator. If this term would be neglected altogether the solution for $v_1(x)$ would be

$$v_1(x) = \exp\left[\frac{i}{\hbar} \int_0^x p(x')\, dx' \right] v_1(0) \tag{2}$$

(and the same for $v_2(x)$ with $-i$ instead of i). However, this is not a good approximation because the neglected term is of order 1. It can be improved with little additional labor by including the diagonal elements; the result is the familiar WKB approximation

$$v_1(x) = \frac{1}{\sqrt{p(x)}} \exp\left[\frac{i}{\hbar} \int_0^x p(x')\, dx' \right] v_1(0) \tag{3}$$

(and the same for $v_2(x)$ with $-i$ instead of i).

The non-diagonal terms in (1) represent interactions between the modulated waves (3). It is true that they are ostensibly of order 1 since they do not contain a factor $\hbar$. Actually, however, they add to the solution only corrections which are of order $\hbar$. In fact, the next correction to $v_1(x)$ is easily seen to involve the integral

$$\int_0^x \frac{p'(x')}{2p(x')} \exp\left[-\frac{2i}{\hbar} \int_0^{x'} p(x'')\, dx'' \right] dx'. \tag{4}$$

This integral is small owing to the rapidly oscillating exponential factor in the integrand. Asymptotic evaluation of this integral by means of the stationary phase method immediately shows that it is of order $\hbar$. Thus *the contribution of the non-diagonal elements is one order of $\hbar$ higher than is indicated by the explicitly appearing factors $\hbar$.* The same remark does not apply

to the diagonal terms, which is the reason why these *must* be included, so as to obtain (3) rather than (2).

It is possible to take the non-diagonal terms in (1) into account by the usual iterative method of perturbation theory, as was done by Bremmer[2]) and others[3]). The result is a series, in which each term contains a factor $\hbar$ more than the previous one, but also an additional integration. This suggests that this perturbation expansion is an expansion in powers of $\hbar L$, where L is the length of the interval under consideration. Actually the integrals again involve rapidly oscillating factors and are therefore very small. In fact, we shall find in sec. 5 that the reflection by a smoothly varying potential barrier is of infinite order in $\hbar$.

2. *The next approximation.* In the proposed method the non-diagonal terms in (1) are not taken into account by means of a series expansion. Rather, we write (1) in the form

$$\frac{dv}{dx} = \frac{1}{\hbar} B(x) v$$

and again apply a transformation,

$$v = Qw,$$

which is chosen such as to diagonalize B

$$Q^{-1}BQ = M, \qquad M = \text{diagonal}. \tag{5}$$

The resulting equation for w will then be

$$\frac{dw}{dt} = \frac{1}{\hbar} Mw - Q^{-1}Q'w, \tag{6}$$

in complete analogy with (1).

In order to show that (6) actually leads to a better approximation than (1), we note that B is diagonal apart from terms of order $\hbar$. Hence Q differs from the unit matrix only by terms of order $\hbar$*),

$$Q = 1 + \hbar Q_1.$$

Consequently the matrix

$$Q^{-1}Q' = \hbar Q^{-1}Q_1' = \hbar Q_1' + O(\hbar)^2$$

is of order $\hbar$, which is one order higher than the last term in (1).

The matrix B has the form

$$B(x) = \begin{pmatrix} ip - \hbar p'/2p & \hbar p'/2p \\ \hbar p'/2p & -ip - \hbar p'/2p \end{pmatrix}.$$

---

*) Strictly speaking this needs not be true since Q is not uniquely determined by (5): but Q *may be so chosen* as to differ from unity only in order $\hbar$.

It is convenient to use temporarily the abbreviation

$$\frac{\hbar p'}{2p^2} = \sin \alpha,$$

so that $\alpha(x)$ is a small quantity of order $\hbar$. Then the diagonalized matrix M is found to be

$$M = \begin{pmatrix} ip \; e^{i\alpha} & 0 \\ 0 & -ip \; e^{-i\alpha} \end{pmatrix}$$

and the diagonalizing matrix Q is

$$Q = \begin{pmatrix} \cos \tfrac{1}{2}\alpha & i \sin \tfrac{1}{2}\alpha \\ -i \sin \tfrac{1}{2}\alpha & \cos \tfrac{1}{2}\alpha \end{pmatrix}. \tag{7}$$

It follows that

$$Q^{-1}Q' = \frac{-\alpha'}{2 \cos \alpha} \begin{pmatrix} \sin \alpha & -i \\ i & \sin \alpha \end{pmatrix},$$

so that the equation (6) for w becomes

$$\frac{d}{dx} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \frac{1}{\hbar} \begin{pmatrix} ip \; e^{i\alpha} + \tfrac{1}{2}\hbar\alpha' \tan \alpha & -\tfrac{1}{2}i\hbar\alpha' \sec \alpha \\ \tfrac{1}{2}i\hbar\alpha' \sec \alpha & -ip \; e^{-i\alpha} + \tfrac{1}{2}\hbar\alpha' \tan \alpha \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

Suppose one stops the process after this step and determines the solution in this approximation. Omitting again the non-diagonal terms, which represent the interaction between w_1 and w_2, one finds

$$w_1(x) = w_1(0) \exp \int_0^x \left\{ \frac{ip}{\hbar} \; e^{i\alpha} + \tfrac{1}{2}\alpha' \tan \alpha \right\} dx$$

$$= w_1(0) \left(p^2 - \frac{\hbar^2 p'^2}{4p^2} \right)^{-\frac{1}{4}} \exp \frac{i}{\hbar} \int_0^x \left(p^2 - \frac{\hbar^2 p'^2}{4p^2} \right)^{\frac{1}{2}} dx.$$

For $w_2(x)$ one finds the same expression with $-i$ instead of i. It appears that this first correction to the WKB approximation has three effects.

(i) It modifies the local wave vector:

$$p \to p \sqrt{1 - \frac{\hbar^2 p'^2}{4p^4}}.$$

The local amplitude is modified accordingly*).

(ii) The components v_1 and v_2 are recombined in a definite way, determined by the transformation matrix Q given by (7).

---

*) Using a different method Bailey [4] found the same modification of the wave number, but not the corresponding modification of the amplitude.

(iii) The non-diagonal terms, which were neglected, contain a factor $\hbar$, but are of order $\hbar^2$ since they also contain an integral with a rapidly oscillating integrand, similar to (4).

3. *The general method.* It is clear that instead of discontinuing the approximation process after (6) one may continue by diagonalizing the matrix in (6) and transforming the vector w accordingly. In this way one can continue step by step, as long as the derivatives exist. After the m-th step the equation has the form

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \left(\frac{1}{\hbar}\,\Omega - Z^{-1}Z'\right)z, \tag{8}$$

where Ω is diagonal and $Z^{-1}Z'$ is of order $\hbar^{m-1}$. Each step transforms away a part of the interaction between the components and incorporates it in the modulation of the two waves. Suppose one stops after the m-th step and determines the solution in the approximation obtained. The result is

$$z_1(x) = \exp\left[\int_0^x \left\{\frac{1}{\hbar}\,\Omega_1(x') - Z'_{11}(x')\right\}\mathrm{d}x'\right]z_1(0). \tag{9}$$

Again the diagonal term has been included in the exponent, but the factor Z^{-1} has been omitted since its difference from unity is of higher order. The non-diagonal terms in (8) have been neglected; at first sight this gives rise to an error of order $\hbar^{m-1}L$, but actually the error is only of order $\hbar^m$, owing to the rapid oscillations of the function which has to be integrated to find the interaction between both components.

Note that the separate steps do not require any integration; only in the final result (9) an integral occurs. The successive transformations of $u(x)$ are determined by the *local* behaviour of the potential, i.e., by V and its derivatives *at the same point* x. The calculation will be done more explicitly in the next section, but we here want to add some comments.

Each step in the process introduces a next higher derivative. It is easily seen that the matrix elements of P are polynomials of the matrix elements of A and of the eigenvalues of A. The eigenvalues of A are themselves algebraic functions of the elements of A and no branch points occur in the region considered, owing to the condition $p^2 > 0$. Hence $P(x)$ and $A(x)$ have as many derivatives as $A(x)$. In this way one sees that the successive transformations can be carried through the m-th step, when A has m derivatives. When A has derivatives of all orders the process can be continued indefinitely. The resulting series of transformations will not in general converge, but constitutes an asymptotic expansion (in the sense of Poincaré).

The transformation matrix P is not uniquely defined by the requirement that it diagonalizes A. For let Γ be any matrix that commutes with A,

which means that $\varGamma$ is diagonal. Then $P\varGamma = \tilde{P}$ also transforms A to the diagonal form $\varLambda$. Setting $u = \tilde{P}\tilde{v}$ one obtains an alternative transformed vector $\tilde{v}$, which is related to the previously used v by

$$ v = \varGamma\tilde{v}, \quad \text{or} \quad v_\nu = \varGamma_\nu\tilde{v}_\nu. \qquad (\nu = 1, 2) $$

Thus this transformation multiplies the separate components with arbitrary factors, but does not mutually combine them. Yet it is not just a change of normalization, because the arbitrary factors $\varGamma_\nu$ may depend on x. It follows that the decomposition into separate components v_ν is not unique and has therefore no physical meaning, as emphasized by Groenewold[5]). Nevertheless, all physical consequences are of course unique, since they can be expressed in the original vector u itself.

The equation governing the time-dependence of $\tilde{v}$ is

$$ \frac{d\tilde{v}}{dx} = \frac{1}{\hbar}\varLambda\tilde{v} - \tilde{P}^{-1}\tilde{P}'\tilde{v} = \left\{ \frac{1}{\hbar}\varLambda - \varGamma^{-1}P^{-1}P'\varGamma - \varGamma^{-1}\varGamma' \right\}\tilde{v}. $$

This shows that the arbitrary diagonal matrix $\varGamma$ can be utilized to cancel the diagonal elements of $P^{-1}P'$; one merely has to find the functions $\varGamma_\nu(x)$ from

$$ \varGamma_\nu' = -(P^{-1}P')_{\nu\nu}\,\varGamma_\nu. $$

This would have the effect that the correction which we have previously added to (2) to obtain (3) is now incorporated in the definition of $\tilde{v}$. Yet we need not make use of this possibility, because on the present method this correction is automatically taken care of in the next step of the approximation.

4. *Higher corrections.* In order to study the higher approximations we first rewrite the work in sec. 2 in a more sophisticated way. With the aid of the Pauli matrices

$$ \sigma_\xi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_\eta = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_\zeta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} $$

one may write (1) in the form

$$ \frac{dv}{dx} = \frac{ip}{\hbar}\sigma_\zeta v + \frac{p'}{2p}\sigma_\xi v - \frac{p'}{2p}v. $$

The last term can be removed by setting

$$ v(x) = p^{-\frac{1}{2}}\tilde{v}(x), $$

which is the transformation described in sec. 3. Then

$$ \frac{d\tilde{v}}{dx} = \left(\frac{ip}{\hbar}\sigma_\zeta + \frac{p'}{2p}\sigma_\xi \right)\tilde{v}. \qquad (10) $$

The diagonalisation of the matrix now amounts to a (complex) rotation in the (ξ, ζ) plane such that the new ζ axis lies along the vector with components $(p'/2p, 0, ip/\hbar)$. To prepare for this transformation we write (10) in the form

$$\frac{d\tilde{v}}{dx} = \sqrt{\frac{p^2}{\hbar^2} - \frac{p'^2}{4p^2}} \, (i\sigma_\zeta \cosh \gamma_1 + \sigma_\xi \sinh \gamma_1) \, \tilde{v},$$

where γ_1 is determined by

$$\tanh \gamma_1 = \frac{\hbar p'}{2p^2}$$

and is a small quantity of order $\hbar$. The square root is absorbed in the coordinate x by defining a new variable

$$\tau_1 = \int_0^x \sqrt{p^2 - \frac{\hbar^2 p'^2}{4p^2}} \, dx.$$

Then (10) takes the form

$$\frac{d\tilde{v}}{d\tau_1} = \frac{1}{\hbar} \, (i\sigma_\zeta \cosh \gamma_1 + \sigma_\xi \sinh \gamma_1) \tilde{v}. \tag{11}$$

The diagonalisation is now achieved by the transformation

$$\tilde{v} = e^{-\frac{1}{2}\gamma_1 \sigma_\eta} \, w;$$

the resulting equation for w is

$$\frac{dw}{d\tau_1} = \left(\frac{i}{\hbar} \sigma_\zeta + \tfrac{1}{2} \frac{d\gamma_1}{d\tau_1} \sigma_\eta \right) w.$$

The idea is to cast this equation in the same form as (11). First we change σ_η into σ_ξ by means of a rotation over $\frac{1}{2}\pi$ about the ζ axis,

$$w = e^{-\frac{1}{4}\pi i \sigma_\zeta} \, \tilde{w},$$

which only changes the phases of both components. Then $\tilde{w}$ obeys the equation

$$\frac{d\tilde{w}}{d\tau_1} = \frac{1}{\hbar} \left(i\sigma_\zeta + \frac{\hbar}{2} \frac{d\gamma_1}{d\tau_1} \sigma_\xi \right) \tilde{w}.$$

Next we define γ_2 and τ_2 by

$$\tanh \gamma_2 = \frac{\hbar}{2} \frac{d\gamma_1}{d\tau_1}, \quad \tau_2 = \int_0^{\tau_1} \sqrt{1 - \frac{\hbar^2}{4} \left(\frac{d\gamma_1}{d\tau_1} \right)^2} \, d\tau_1. \tag{12}$$

Note that γ_2 is of order $\hbar^2$. The equation for $\tilde{w}$ may then be written

$$\frac{d\tilde{w}}{d\tau_2} = \frac{1}{\hbar} \, (i\sigma_\zeta \cosh \gamma_2 + \sigma_\xi \sinh \gamma_2) \, \tilde{w}.$$

This differs from (11) only by the fact that the function $\gamma_1(\tau_1)$ has been replaced with $\gamma_2(\tau_2)$.

Clearly the process can now be repeated step by step, and we only have to study the relation between the successive pairs of quantities γ_m, τ_m. The recursion relations expressing γ_{m+1} and τ_{m+1} in terms of γ_m and τ_m are according to (12)

$$\tanh \gamma_{m+1} = \frac{\hbar}{2}\frac{d\gamma_m}{d\tau_m}, \qquad \tau_{m+1} = \int\limits_0^{\tau_m} \sqrt{1 - \frac{\hbar^2}{4}\left(\frac{d\gamma_m}{d\tau_m}\right)^2}\, d\tau_m. \qquad (13)$$

Note that these relations are valid even for $m = 0$, if one defines

$$\gamma_0 = \log p, \quad \tau_0 = \int\limits_0^x p\, dx.$$

It appears that each step has the same three effects as mentioned in sec. 2, viz.,

(i) It modifies the local wave number according to (13).

(ii) The two components are recombined by the transformation matrix

$$e^{-\frac{1}{2}\pi i \sigma_\xi}\, e^{-\frac{1}{2}\gamma_m \sigma_\eta}.$$

(iii) An additional factor $\hbar$ appears in the non-diagonal terms that describe the remaining interaction between both components.

5. *The reflection problem.* As mentioned before there is no unique way of decomposing the solution into a wave travelling to the right and a wave to the left, unless the potential is a constant. In order to define reflection and transmission by a potential barrier we therefore suppose

$$p(x) = \text{const.} = p_0 > 0 \quad \text{for} \quad x < 0;$$
$$p(x) = \text{const.} = p_1 > 0 \quad \text{for} \quad x > x_1.$$

In these regions one has

$$\gamma_0 = \text{const.}, \quad \gamma_1 = \gamma_2 = \ldots - 0;$$
$$\tau_0 = \tau_1 = \tau_2 = \ldots = p_0 x \quad \text{or} \quad p_1 x \text{ respectively.}$$

The exact solution in these regions is therefore found after the first step:

$$\left.\begin{array}{l} v_1(x) = a\, e^{ip_0 x/\hbar} \\ v_2(x) = b\, e^{-ip_0 x/\hbar} \end{array}\right\} \text{ for } x < 0 \qquad (14)$$

$$\left.\begin{array}{l} v_1(x) = c\, e^{ip_1 x/\hbar} \\ v_2(x) = 0 \end{array}\right\} \text{ for } x > x_1 \qquad (15)$$

The last line imposes the condition that there is no incident wave from the right (see figure).
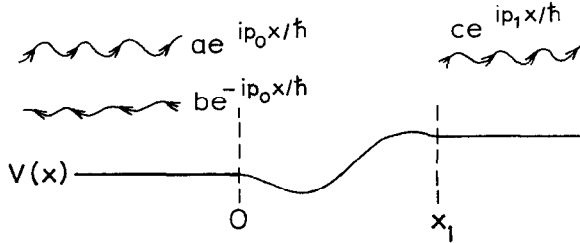
Fig. 1

Reflection and transmission are determined by b/a and c/a. More precisely, the number of particles passing through any point x per unit time is giben by the probability current

$$S = \frac{\hbar}{2im}\left(\varphi^* \frac{d\varphi}{dx} - \varphi \frac{d\varphi^*}{dx}\right) = \frac{1}{2im}(u_1^* u_2 - u_1 u_2^*) = \frac{p}{m}(|v_1|^2 - |v_2|^2).$$

For $x > x_1$ this is equal to $S_1 = (p_1/m)\,|c|^2$, for $x < 0$ it is $S_0 = (p_0/m)(|a|^2 - |b|^2)$. Hence the reflection and transmission coefficients are

$$R = |b/a|^2, \qquad T = (p_1/p_0)\,|c/a|^2.$$

Now take the m-th order solution $(z_1(x), z_2(x))$ for $-\infty < x < \infty$, specified by the boundary condition at $x = x_1$:

$$z_1(x_1) = c_1\, e^{ip_1 x_1/\hbar}, \quad z_2(x_1) = 0.$$

For $x > x_1$ it coincides with the desired solution (15). For $x < x_1$ the functions $z_1(x)$ and $z_2(x)$ can be found by the method in the previous section. Without explicit calculation, however, it is already clear that $z_2(x) = O(\hbar^m)$ for all x. Consequently for $x < 0$

$$v_2(x) = z_2(x) = O(\hbar^m). \quad (x < 0)$$

Hence $b = O(\hbar^m)$ so that $R = O(\hbar^{2m})$. It follows that *when $p(x)$ does not vanish and has derivatives of all orders in $-\infty < x < \infty$ (including $x = 0$ and $x = x_1$) the reflection is small of infinite order in $\hbar$.*

It also follows that T differs from unity only by terms of infinite order in $\hbar$. The main effect of the potential on observations in the external regions is therefore the phase shift due to the modification of the optical path. One has

$$\sqrt{p_1}z_1(x_1) = \sqrt{p_0}\exp[i\tau_m(x_1)/\hbar]\,z_1(0),$$

so that the phase shift is given by

$$\frac{\sqrt{p_1}c}{\sqrt{p_0}a} = \exp\frac{i}{\hbar}\{\tau_m(x_1) - p_1 x_1\} + O(\hbar^\infty).$$

Another consequence is that the energy densities on both sides of the

barrier are equal, apart from terms of order $\hbar^\infty$. The energy density is

$$\varepsilon = \frac{\hbar^2}{2m}\left|\frac{d\varphi}{dx}\right|^2 + V|\varphi|^2 = E(|v_1|^2 + |v_2|^2) + \left(V - \frac{p^2}{2m}\right)(v_1^* v_2 + v_1 v_2^*).$$

The last term oscillates rapidly, and may therefore be omitted to obtain $\bar{\varepsilon}$, the energy density averaged over space. Since in the external regions v_2 is of order $\hbar^\infty$, one finds

$$\bar{\varepsilon}(x < 0) = E|a|^2, \quad \bar{\varepsilon}(x > x_1) = E|c|^2.$$

The fact that $T = 1$ may be expressed by

$$p_0 \bar{\varepsilon}(x < 0) = p_1 \bar{\varepsilon}(x > x_1) + O(\hbar^\infty).$$

This is simular to Kulsrud's result[6] concerning the adiabatic invariance of the quantity E/ω for an harmonic oscillator with slowly varying frequency ω and with energy E. The reason why in our case one has to *multiply* the energy by p rather than divide by it, is that the expression for the energy of a Schrödinger particle is different than for the classical harmonic oscillator.

In the case of an electromagnetic wave in a layered medium the same equations hold, when the refractive index is identified with $p(x)$, and the frequency ω is substituted for $\hbar^{-1}$. Hence the reflection by a slab of such a medium is of order $\omega^{-\infty}$, provided that the refractive index is an infinitely differentiable function of x (no discontinuous boundaries!) and nowhere zero. The energy density is $\varepsilon = p^2(|v_1|^2 + |v_2|^2)$, so that the adiabatic invariant is ε/p. However, these statements presuppose that the refractive index is independent of ω, which is rather unrealistic.

REFERENCES

1) See also Heading, J., An Introduction to Phase-Integral Methods (Methuen, London, 1962).
2) Bremmer, H., Physica **15** (1949) 593; Handb. der Physik **16** (S. Flügge ed., Springer-Verlag, Berlin, 1953) p. 552.
3) Fröman, N. and Fröman, P. O., JWKB Approximation (North-Holland Publ. Cy., Amsterdam, 1965).
4) Bailey, V. A., Phys. Rev. **96** (1954) 865.
5) Groenewold, H. J., Mat. Fys. Medd. **30** (1956) No. 19.
6) Kulsrud, R. M., Phys. Rev. **106** (1954) 865.