Annals of Mathematics, 120 (1984), 89–118

# Semiclassical analysis of low lying eigenvalues, II. Tunneling*

By Barry Simon


## Abstract

We discuss the leading asymptotics of eigenvalue splittings of $-\frac{1}{2}\Delta + \lambda^2 V$ in the limit as $\lambda \to \infty$, and where $V$ is a non-negative potential with several zeros. For example, if $E_0(\lambda)$, $E_1(\lambda)$ are the two lowest eigenvalues in a situation where $V$ has precisely two zeros, $a$ and $b$, related by a symmetry, then $\lim_{\lambda \to \infty} -(\lambda)^{-1} \ln[E_1(\lambda) - E_0(\lambda)]$ is given as the distance from $a$ to $b$ in a certain Riemann metric.


## 1. Introduction

From the earliest days of quantum mechanics, it has been clear that a basic difference from classical mechanics concerns the ability of particles to tunnel between two regions separated by a classically forbidden region. Quantitative estimates of this phenomenon first became important in work on lifetimes for $\alpha$-decays and for widths of Stark lines. Mathematical analysis of these ideas has always been made difficult by the fact that a precise definition of lifetime is not easy, and indeed, there is no very good reconciliation of the fact that pure exponential decay rates are observed to incredible accuracy in nuclear decays while mathematically pure exponential decay is forbidden in systems whose energy is bounded below. The modern theory of complex scaling [4], [9], [44] (see [40], [41] for reviews) has provided a precise meaning to lifetimes which can then be mathematically analyzed, although the fact that this definition is not a time dependent one has led to some controversy about whether the results of this analysis are really mathematical justifications of tunneling calculations. Not all complex scaling results concern conventional tunneling, but Herbst's extension of complex scaling to Stark problems [32] does allow one to analyze some conventional tunneling situations.

*Research partially supported by USNSF grant MCS-81-20833

Another situation where tunneling is relevant, and which is in many ways cleaner than lifetime calculations, concerns multiwell problems. To describe this class of problems, consider the very simplest example:

$$\frac{-d^2}{dx^2} + x^2 - 2\beta x^3 + \beta^2 x^4 = D(\beta). \tag{1.1}$$

$D(\beta)$ has purely discrete spectrum for $\beta > 0$. Because of the symmetry of the potential

$$x^2 - 2\beta x^3 + \beta^2 x^4 = \beta^2 x^2 (x - \beta^{-1})^2$$

under the map $x \mapsto \beta^{-1} - x$ of reflection in $\beta^{-1}/2$, we have two identical wells about $x = 0$ and $x = \beta^{-1}$. Classically, if $\beta$ is small and we look at the behavior of the system for energies near 1, the two wells are completely decoupled. In the quantum systems, the two wells are coupled by tunneling and the degeneracy of the lowest eigenvalues in each well is removed by this coupling. The size of the gap, for $\beta$ large, is exponentially small, and the various constants in the asymptotics are determined by tunneling considerations. Since the splitting of eigenvalues is such a simple object, it is susceptible to precise analysis. Moreover, there is a direct link with time dependent phenomena: If $\Omega_0, \Omega_1$ are the two lowest eigenvectors suitably normalized, then $\Omega_0 + \Omega_1$ (resp $\Omega_0 - \Omega_1$) are concentrated primarily in the well at $x = 0$ (resp at $x = \beta^{-1}$). Thus, since

$$e^{-itH}(\Omega_0 + \Omega_1) = e^{-itE_0}\left(\Omega_0 + e^{-it(E_1 - E_0)}\Omega_1\right),$$

we see that $\pi/(E_1 - E_0)$ is precisely the time needed to evolve from a state concentrated primarily in one well into one primarily in the other.

Our main goal in this paper will be the determination of the leading asymptotics of quantities like $E_1 - E_0$ as $\beta \to 0$, especially in multidimensional analogs of (1.1). Rigorous asymptotics of tunneling parameters have a relatively brief history. The pioneering work of Titchmarsh [53] on decay, and Kac-Thompson [34] on double wells, established primarily exponentially small upper bounds without attention to precise constants. With two exceptions to be noted below, the more recent results on precise rigorous asymptotics is restricted to one dimensional problems, often relying on ODE techniques. It was for eigenvalue splitting problems that these results were first obtained for successively more complex problems by Harrell in a series of papers [28], [29], [30]. More recently, Combes, Duclos and Seiler [17] and Jona-Lasinio, Martinelli and Scoppola [33] have obtained interesting further results in this direction, and Davies [19], [20], [21], [22] has considered an abstract framework for some of the related phenomena. For decay problems, the first results were obtained by Harrell-Simon [31] (who considered the Stark problem for hydrogen, which, while 3-dimensional, was studied by separating in a suitable coordinate system to essentially one dimensional systems), with recent results by Ashbaugh-Harrell [7], Ashbaugh-Sundberg [8] and Corngold et al. [18]. It happens that large order asymptotics of divergent perturbation series are related to tunneling (see Bender-Wu [10] or the review by Simon [48]), and for the so-called anharmonic oscillator series, leading asymptotics were obtained by Harrell-Simon [31] using tunneling ideas and ODE methods. For our purposes here, a more useful approach to this problem uses path integrals and was developed by Simon [50], Spencer [52] and Breen [11] in successively more detail. In principle, these works using path integral methods are not restricted to one dimension.

I know of two previous works involving multidimensional tunneling. There is one paper by Harrell [30] on double wells, but his results are "essentially one dimensional" in that in our language below, his problems are chosen so that the minimizing geodesic is a straight line. As noted above, the path integral approach to large order is capable of working in higher dimensions and, indeed, Breen [12] has a result in infinitely many variables (a spatially cutoff quantum field theory).

Our technique in this paper has some elements in common with that of Breen [12]. We use path integrals, specifically the method of large deviations. It is fortunate that just before I began thinking about these problems, I heard some beautiful lectures by Varadhan [56] on large deviations.

Let us describe a simple situation we want to analyze here. Our analysis of this problem will appear in Sections 2–4, extensions in Sections 5–7; a sketch of the arguments appeared in [47].

Let $V$ be a function on $R^\nu$ obeying (a fourth hypothesis appears later):

(1) $V$ is $C^\infty$ and non-negative;

(2) $V$ is strictly positive at $\infty$; i.e., for some $R$, $\varepsilon > 0$, $V(x) > \varepsilon$ if $|x| \geq R$;

(3) $V$ vanishes at exactly two points $a, b$ and $\partial^2 V/\partial x_i\, \partial x_j(x)$ is a non-singular matrix for $x = a, b$.

As we will see later, the smoothness of $V$ and the non-degeneracy of the minima are not really important. Positivity near $\infty$ is critical as well as the fact that zeros can be divided into two or more disjoint sets. We are interested in studying the operator

$$H(\lambda) = -\tfrac{1}{2}\Delta + \lambda^2 V(x) \tag{1.2}$$

for $\lambda$ large. Parenthetically, we note that while (1.1) does not have this form, it is related to a suitable $H(\lambda)$ where $V(x) = x^2 - 2x^3 + x^4$ and $\lambda = \beta^{-2}$. For if $U$ is the unitary operator with

$$UxU^{-1} = \beta^{-1}x;\ U\frac{d}{dx}U^{-1} = \beta\frac{d}{dx},$$

then

$$\beta^{-2}UHU^{-1} = \frac{-d^2}{dx^2} + \beta^{-4}(x^2 - 2x^3 + x^4).$$

We are interested in very fine asymptotics of eigenvalue differences for this problem. Crude asymptotics on individual eigenvalues will be important. These latter asymptotic results are well-known folk theorems in the physics literature: Surprisingly, only recently have rigorous proofs been written down. See [46] for the results we state below; we note that these theorems can also be proved using the methods of Davies [23]; we should also mention the work of Combes [15], Combes et al. [16] and Reed-Simon [40] on the one dimensional case.

Let $\{\omega_i^{(a)}\}_{i=1}^{\nu}$ and $\{\omega_i^{(b)}\}_{i=1}^{\nu}$ be positive numbers so that $\frac{1}{2}[\omega_i^{(\#)}]^2$ are the eigenvalues of $\frac{1}{2}(\partial^2 V/\partial x_i\, \partial x_j)(\#)$ for $\# = a, b$. Consider the union of the two "sets with multiplicities," $\sum_{i=1}^{\nu}(n_i + \frac{1}{2})\omega_i^{(\#)}$ where $n_i = 0, 1, 2, \ldots$ and let $e_0 \leq e_1 \leq e_2 \leq \cdots$ be a listing of these sets labeled in increasing order. Then:

THEOREM 1.1 (*proven in* [46]). *Let $H(\lambda)$ be given by* (1.2) *where $V$ obeys* (1)–(3). *Then for each $n$, when $\lambda$ is sufficiently large, $H(\lambda)$ has at least $n + 1$ eigenvalues, $E_0(\lambda), \ldots, E_n(\lambda), \ldots$ and*

$$\lim_{j \to \infty} E_j(\lambda)/\lambda = e_j. \tag{1.3}$$

Intuitively, this result comes from the following: When $\lambda$ is large for eigenvalues not to be $O(\lambda^2)$, the corresponding eigenvectors must live very near either $a$ or $b$. Near $a$, $H(\lambda)$ looks like a sum of harmonic oscillators of frequencies $\lambda\omega_i^{(a)}$ and similarly at $b$. The "union" of the two sets of oscillators has eigenvalue precisely $\lambda e_j$.

In [46], one finds asymptotic series to all orders (in $\lambda^{-1}$ and $\lambda^{-1/2}$ respectively) for the $E_j$ and for the corresponding eigenvectors $\Omega_j(\lambda)$. One consequence of these series is the following: Let $j_a$, $j_b$ be functions supported in very small neighborhoods of $a$ and $b$. Then either there is a rapid eigenvalue degeneracy or the eigenfunctions live in a single well in the following sense (Cor. 5.4 of [46]):

THEOREM 1.2. *Let $H(\lambda)$ be given by* (1.2) *where $V$ obeys* (1)–(3), *and suppose $j$ is such that $E_j(\lambda)$ is non-degenerate for all large $\lambda$ (although perhaps not at $\lambda = \infty$, i.e., $e_j$ may be degenerate). Then one of the following holds*:

(a) $\|j_a\Omega_j(\lambda)\| = O(\lambda^{-N})$ *for all $N$ or*

(b) $\|j_b\Omega_j(\lambda)\| = O(\lambda^{-N})$ *for all $N$ or*

(c) *There is another eigenvalue $E'(\lambda)$ with $|E'(\lambda) - E(\lambda)| = O(\lambda^{-N})$ for all $N$.*

One of our goals here will be to prove the following (strengthening of Thm. 1.2) in Sections 2, 3:

THEOREM 1.3. *There is a* $C > 0$ *so that under the hypotheses of Theorem* 1.2, *one of the following holds*:

(a) $\overline{\lim}\left[-\frac{1}{\lambda}\ln\|j_a\Omega_j(\lambda)\|\right] > C$ *or*

(b) $\overline{\lim}\left[-\frac{1}{\lambda}\ln\|j_b\Omega_j(\lambda)\|\right] > C$ *or*

(c) *There is another eigenvalue* $E'$ *with* $\underline{\lim}\ -\frac{1}{\lambda}\ln|E'(\lambda) - E(\lambda)| > C$.

This result, since no value of $C$ is involved, will not be very hard to prove. More interesting and subtle are results which evaluate the constant in $|E' - E|$. In the rest of this section, we discuss the ground state (lowest eigenvalue) where we will get upper *and* lower bounds on the difference and determine the exact asymptotics; in Section 6, we discuss excited states. We will suppose that the ground state has a piece in both wells so that the second eigenvalue is exponentially close to the lowest, i.e., we assume:

(4) $\underline{\lim}_{\lambda \uparrow \infty}\ (\|j_a\Omega_0(\lambda)\|\ \|j_b\Omega_0(\lambda)\|) > 0.$

Actually, if the product is bounded from below by $\lambda^{-k}$ for any $k$, that would suffice for our considerations. There is one case where (4) always holds: For suppose that $R$ is a Euclidean map from $R^\nu$ to itself and let $(U\varphi)(x) = \varphi(Rx)$. Suppose that $V(Rx) = V(x)$ for some $R$ with $Ra = b$. Then $UH(\lambda)U^{-1} = H(\lambda)$ and so $(U\Omega_0) = \Omega_0$ which implies that $\lim\|j_a\Omega_0\| = \lim\|j_b\Omega_0\| = \frac{1}{2}$. In this case with symmetry, (4) is thus automatic. In case $R$ is a reflection such as occurs in the double well, (1.1), the geodesic geometry below is simplified; e.g., the geodesic bisector is a plane. But if $R$ is a rotation, e.g., $V(x, y) = (x-1)^2(x+1)^2 + (xy-1)^2$; $R(x, y) = (-x, -y)$, then the geodesic bisector is not a hyperplane.

To state our main result for this basic situation we need to introduce a metric discussed initially by Agmon [1], [2] in a related context (see below); it is very close to a metric used by Jacobi in his studies of classical mechanics.

*Definition*. Given a function $V(x)$ obeying (1-3), we define the *Agmon metric*, $\rho$, by

$$\rho(x, y) = \inf\left\{\int_0^1 \sqrt{2V(\gamma(s))}\ |\dot{\gamma}(s)|\, ds \,\middle|\, \gamma(0) = x, \gamma(1) = y\right\}, \tag{1.4}$$

the geodesic distance in the Riemann metric $2V(x)\,dx^2$ conformal to the Euclidean metric.

When the Agmon metric appears in the proof in Section 3, it will not be in the form (1.4) but in an equivalent form found by Carmona-Simon [13]:

PROPOSITION 1.4 ([13]). *Let* $\rho$ *be given by* (1.4). *Then*

$$(1.5)\quad \rho(x,y) = \inf_{\gamma,T}\left\{\frac{1}{2}\int_0^T|\mathring{\gamma}(s)|^2\,ds + \int_0^T V(\gamma(s))\,ds \,\middle|\, \gamma(0)=x, \gamma(T)=y\right\},$$

where we minimize over T also.

*Sketch of proof.* Let $\tilde{\rho}$ denote the right side of (1.5). Since $ab \le \frac{1}{2}(a^2+b^2)$, we can take $a = \sqrt{2V(\gamma(s))}$; $b = |\mathring{\gamma}(s)|$ and see that $\rho \le \tilde{\rho}$ (we use the fact that arc length is invariant under parameterization). Conversely, given any trial path $\gamma$ for $\rho$, we reparametrize it so that $|\mathring{\gamma}| = \sqrt{2V(\gamma(s))}$ (see below) and use this new path as a trial function for $\tilde{\rho}$ and so that $\tilde{\rho} \le \rho$. □

At points where $V$ vanishes, one may not be able to reparametrize $\gamma$ so that the above holds and still arrange that $T$ be finite. Thus, before reparametrizing, we shift the path with a small change of arc length so that zeros of $V$ are avoided. If a zero of $V$ is an endpoint, we neglect to reparametrize in a very small neighborhood of that end point. These considerations of zeros have an important aspect: The minimum problem (1.4) always has a minimizing path. If $x$ and $y$ are zeros of $V$, then (1.5) may not possess a minimizing path if $T < \infty$ is required. If both endpoints are zeros, there is a path parameterized by $(-\infty, \infty)$ so that $\lim_{T\to-\infty}\gamma(s) = x$, $\lim_{T\to\infty}\gamma(s) = y$. This minimizing path for (1.5) with $x = a$, $y = b$ (run from $-\infty$ to $+\infty$) is called an *instanton*; we will discuss this further below.

We remark that the quantity (1.5) is just the classical mechanical action, for a particle moving in a potential $-V$ (note change of sign). The above remarks about infinite times are a reflection of the well known fact that a particle rolling uphill to stop at an unstable equilibrium takes infinitely long to get there.

Our main result is

THEOREM 1.5. *Let* $V$ *be a function on* $R^\nu$ *obeying* (1)–(4). *Let* $H(\lambda) = -\frac{1}{2}\Delta + \lambda^2V(x)$ *and let* $E_1(\lambda), E_0(\lambda)$ *be the two lowest eigenvalues of* $H(\lambda)$. *Then*

$$(1.6)\qquad \lim_{\lambda\to\infty} -\lambda^{-1}\ln[E_1(\lambda) - E_0(\lambda)] = \rho(a,b)$$

*where* $\rho(a,b)$ *is the distance from* $a$ *to* $b$ *in the Agmon metric.*

If $\nu = 1$, the Agmon geodesic is a straight line from $a$ to $b$, so that $\rho(a,b) = \int_a^b\sqrt{2V(x)}\,dx$, the "WKB" answer. (1.6) is then (at least for the Hamiltonian (1.1) and related models) a result of Harrell [28], [29]. We emphasize that Harrell obtains more than just the leading behavior; his methods are mainly ODE methods and restricted to one dimension.

As already mentioned, the function minimizing (1.5) for $x = a, y = b$ (and parameterized to run from $-\infty$ to $\infty$) is called an instanton, $\rho(a, b)$ is its "action" and our main theorem says that "tunneling is determined by the action of the instanton." This fact is a standard piece of wisdom from the physics literature (see e.g. [14], [25], [39]); our result is a rigorous justification of these ideas from the physical literature.

Agmon introduced the "Agmon metric" in his study of the decay of $L^2$ solutions of $(-\Delta + V)u = 0$ at infinity (for $V$'s relevant to us here, this metric actually appears first in a paper of Lithner [36]). Its appearance here suggests that the asymptotics of $E_1 - E_0$ will be connected to exponential decay of eigenfunctions of $H(\lambda)$, an idea due to Harrell [28], [30]. In fact, in Section 2 we reduce the proofs of Theorems 1.3 and 1.5 to results on decay of eigenfunctions. We then give two independent proofs of this decay: In Section 3, we use path integral techniques which was our original proof. In Section 4, we pander to those who dislike path integrals and provide a PDE proof patterned after Agmon's proofs for the large distance problem.

Further results including some on eigenvalue pairs other than the lowest occur in Sections 5–7.

It is a pleasure to thank S. Agmon, E. Davies, C. Fefferman, I. Sigal and S. Varadhan for useful discussions; H. Dym and I. Sigal for the hospitality of the Weizmann Institute, where part of this work was done.

## 2. Reduction to decay of eigenfunctions

In this section, we reduce the proofs of our two new results, Theorems 1.3 (for the ground state) and 1.5 to results on the decay of the ground state of $H(\lambda)$, i.e., to the normalized-$L^2$ vector $\Omega_0(\lambda; x)$ obeying

$$H(\lambda)\Omega_0(\lambda; x) = E_0(\lambda)\Omega_0(\lambda; x). \tag{2.1}$$

The decay results are proved by two distinct methods in the next two sections. Theorem 1.5 requires a rather refined decay estimate; Theorem 1.3 only needs the following rather crude estimate:

THEOREM 2.1. *Let hypotheses* (1), (2), (3) *hold. Then*:

(a) *For any* $\varepsilon > 0$, *there is a* $C_\varepsilon > 0$ *so that for all sufficiently large* $\lambda$, *if* $|x - a| > \varepsilon$ *and* $|x - b| > \varepsilon$, *then*

$$|\Omega_0(\lambda; x)| \leq e^{-C_\varepsilon \lambda}. \tag{2.2}$$

(b) *For some* $R_0$ *and* $D$, *if* $|x| > R_0$ *and* $\lambda$ *is sufficiently large, then*

$$|\Omega_0(\lambda; x)| \leq e^{-D\lambda|x|}. \tag{2.3}$$

We prove this in the next section. We also use the following basic equality whose relevance to this type of problem was emphasized by Kac-Thompson [34]:

PROPOSITION 2.2. *For any* $C^1$ *uniformly bounded function* $f$:

$$(f\Omega_0, (H(\lambda) - E_0(\lambda))f\Omega_0) = \tfrac{1}{2}((\nabla f)\Omega_0, (\nabla f)\Omega_0). \tag{2.4}$$

*Proof.* We note that

$$\begin{aligned}[f, [f, (H(\lambda) - E_0(\lambda))]] &= [f, [f, -\tfrac{1}{2}\Delta]] \\ &= -(\nabla f)^2,\end{aligned}$$

and take expectations of this equality in the vector $\Omega_0$. □

(2.4) says that $H - E_0$ is a "Dirichlet form" and this has been used in a variety of aspects of mathematical quantum mechanics; see e.g. Gross [27], Rosen [42], Albeverio et al. [5], [6] or Glimm-Jaffe [26]. We give the proof of Theorem 1.3 in the ground state case here, and the general proof in Section 6.

*Proof of Theorem* 1.3 (for $E = E_0$) (given Thm. 2.1). Pick a function $g \in C^\infty$ so that $|g| \leq 1$ and $g$ is 1 (resp. $-1$) in a neighborhood of $a$ (resp. $b$) and so that $\nabla g$ is uniformly bounded. Let $\langle g \rangle_\lambda = \int g\Omega_0^2\,dx$ and let $f = g - \langle g \rangle_\lambda$ ($\lambda$ dependent). Since

$$(f\Omega_0, \Omega_0) = 0$$

by construction, we have, by (2.4), that

$$E_1(\lambda) - E_0(\lambda) \leq \tfrac{1}{2}((\nabla f)\Omega_0, (\nabla f)\Omega_0)/(f\Omega_0, f\Omega_0). \tag{2.5a}$$

Moreover, by Theorem 2.1, it is easy to see that

$$((\nabla f)\Omega_0, (\nabla f)\Omega_0) \leq e^{-2C\lambda} \tag{2.5b}$$

for some $C$ and $\lambda$ large. There are now two cases to consider:

*Case* 1. $\overline{\lim} - \dfrac{1}{\lambda}\ln(f\Omega_0, f\Omega_0) \leq C$. In this case, (2.5a) and (2.5b) immediately imply that $\underline{\lim} - \dfrac{1}{\lambda}\ln|E_1 - E_0| \geq C$.

*Case* 2. For some sequence $\lambda_n \to \infty$,

$$(f\Omega_0, f\Omega_0) \leq e^{-\frac{1}{2}C\lambda_n}. \tag{2.6}$$

Suppose that for infinitely many $\lambda_n$, $\langle g \rangle_\lambda \geq 0$. Then $|f| \geq 1$ near $b$ and so (2.6) implies $\|j_b\Omega_0\|^2 \leq e^{-\frac{1}{2}c\lambda_n}$, i.e., $\underline{\lim} - (1/\lambda)\ln\|j_b\Omega_0\| > 0$. If $\langle g \rangle_\lambda \leq 0$, we get information on $\|j_a\Omega_0\|$ instead. □

To get the finer result, Theorem 1.5, we need more information on $\Omega_0$. In the next section, we will prove

THEOREM 2.3. (a) *If hypotheses* (1)–(3) *hold, then for any* $x$

$$\overline{\lim} \frac{1}{\lambda} \ln|\Omega_0(\lambda, x)| \leq -\min(\rho(x,a), \rho(x,b)),$$

*the limit being uniform on compact subsets of* $x$.

(b) *If hypotheses* (1)–(4) *hold, then for any* $x$,

$$\underline{\lim} \frac{1}{\lambda} \ln|\Omega_0(\lambda, x)| \geq -\min(\rho(x,a), \rho(x,b)).$$

We will prove Theorem 1.5 by obtaining upper and lower bounds on $E_1 - E_0$.

*Proof of Theorem* 1.5 (*upper bound*) (given Thms. 2.1 and 2.3). Let

$$d(x) = [\rho(x,a) - \rho(x,b)]/\rho(a,b)$$

and given $\delta$, let $d_\delta(x)$ be a smooth function with $|d - d_\delta| \leq \delta$, obtained for example by convolutions of $d$. Pick $\alpha > 0$ and let $h$ be a $C^\infty$ function on $(-\infty, \infty)$ which is $-1$ on $(-\infty, -\alpha)$ and 1 on $(\alpha, \infty)$. Let $g(x) = H(d_\delta(x))$. Thus $g$ is a smooth function so that $\nabla g$ is supported in a small neighborhood of the geodesic bisector of $a, b$, i.e., in a small neighborhood of $\{x | d(x) = 0\} = \{x | \rho(x,a) = \rho(x,b)\} \equiv B$. By taking $\delta, \alpha$ small, we can arrange that the neighborhood is an arbitrarily small neighborhood. In particular, since

$$\min\{\min(\rho(x,a), \rho(x,b)) | x \in B\} = \tfrac{1}{2}\rho(a,b)$$

we can, given $\varepsilon$, find $\alpha, \delta$ so that

$$\min\{\min(\rho(x,a), \rho(x,b)) | x \in \operatorname{supp} \nabla g\} \geq \tfrac{1}{2}\rho(a,b) - \varepsilon.$$

As in the last proof, let $f = g - \langle g_\lambda \rangle$. Then

$$E_1(\lambda) - E_0(\lambda) \leq \tfrac{1}{2}((\nabla f)\Omega_0, (\nabla f)\Omega_0)/(f\Omega_0, f\Omega_0).$$

As in that proof, if $\underline{\lim}(1/\lambda)\ln(f\Omega_0, f\Omega_0) < 0$, either $\|j_a\Omega_0\|$ or $\|j_b\Omega_0\|$ becomes small, so that hypothesis (4) implies

$$\lim \frac{1}{\lambda} \ln(f\Omega_0, f\Omega_0) = 0.$$

Moreover, by the bound (2.3) (to control large $|x|$) and Theorem 2.3, we have that

$$\begin{aligned}\overline{\lim} \frac{1}{\lambda} \ln((\nabla f)\Omega_0, (\nabla f)\Omega_0) &\leq -2\min(\min(\rho(x,a), \rho(x,b) | x \in \operatorname{supp} \nabla g) \\ &\leq -\rho(a,b) + 2\varepsilon.\end{aligned}$$

Since $\varepsilon$ is arbitrary, we find that

$$\overline{\lim_{\lambda\to\infty}}\ \frac{1}{\lambda}\ln|E_1(\lambda) - E_0(\lambda)| \leq -\rho(a,b)$$

as required. $\square$

For the lower bound, we let $\Omega_1$ be the second lowest eigenfunction normalized so that $\Omega_1(a) > 0$ and $g_\lambda(x) = \Omega_1(x)/\Omega_0(x)$. As a preliminary, we need to know that for $\lambda$ large, there is $C$ so that $g_\lambda(x) \geq C$ for $x$ very near $a$ and $g_\lambda(x) \leq -C$ very near $b$. To prove that, we need:

LEMMA 2.4. *Suppose that $W_\lambda(x) \to W_\infty(x)$ as $\lambda \to \infty$ where the convergence is uniform on compacts for $W_\lambda$ and all its derivatives. Suppose $(-\frac{1}{2}\Delta + W_\lambda)\varphi_\lambda = E_\lambda\varphi_\lambda$, $(-\frac{1}{2}\Delta + W_\infty)\varphi_\infty = E_\infty\varphi_\infty$ (in differential equation sense; $\varphi_\infty$ may not be globally $L^2$), $E_\lambda \to E_\infty$, and $\varphi_\lambda \to \varphi_\infty$ in $L^2_{\text{loc}}$. Then $\varphi_\lambda \to \varphi_\infty$ locally uniformly.*

*Proof.* Let $\psi \in C^\infty$, $\eta \in C_0^\infty$, both real valued. Then

$$\eta|\nabla\psi|^2 + \eta(\psi\Delta\psi) - \tfrac{1}{2}(\Delta\eta)\psi^2 = \nabla\cdot\left(\eta\psi\nabla\psi - \tfrac{1}{2}\psi^2\nabla\eta\right),$$

so that

$$\int\eta|\nabla\psi|^2 = \frac{1}{2}\int(\Delta\eta)\psi^2 - \int\eta(\psi\Delta\psi). \tag{2.7}$$

This equality then shows that if $\psi, \Delta\psi \in L^2_{\text{loc}}$, so is $\nabla\psi$ and if $\psi, \Delta\psi \to 0$ in $L^2_{\text{loc}}$, then $\nabla\psi \to 0$ in $L^2_{\text{loc}}$. By hypotheses on $W$, $\Delta\varphi_\lambda \to \Delta\varphi_\infty$ in $L^2_{\text{loc}}$ so that $\nabla(\varphi_\lambda - \varphi_\infty) \to 0$ in $L^2_{\text{loc}}$. By an inductive argument using the derivatives of the eigenfunction equation, one sees that $D^\alpha\varphi_\lambda \to D^\alpha\varphi_\infty$ in $L^2_{\text{loc}}$ and so by Sobolev estimates in $L^\infty_{\text{loc}}$. $\square$

LEMMA 2.5. *Under hypotheses* (1)–(4), *there exists a $C > 0$ and $\Lambda_0$ so that for $\lambda > \Lambda_0$, and $|x - a| \leq \lambda^{-\frac{1}{2}}$, $g_\lambda(x) > C$ and for $|x - b| \leq \lambda^{-\frac{1}{2}}$, $g_\lambda(x) < -C$.*

*Proof.* We require some of the machinery of [46] on the form of $\Omega_0$ and $\Omega_1$ as $\lambda \to \infty$. Let $\zeta_a(\lambda, x)$, $\zeta_b(\lambda, x)$ be the normalized ground states for the Hamiltonians obtained from $H(\lambda)$ by replacing $V$ by the quadratic Taylor approximation about $a$ and $b$; so

$$\zeta_a(\lambda, x) = \lambda^{\nu/2}\kappa_a\left(\lambda^{\frac{1}{2}}(x - a)\right)$$

for a suitable Gaussian $\kappa_a$ and a similar formula for $\zeta_b$.

The results of [46] say that if $P_\lambda$ is the projection onto the span of $\zeta_a$ and $\zeta_b$, then $\|(1 - P_\lambda)\Omega_j\| \to 0$ for $j = 0, 1$. Thus, there exist positive $\alpha(\lambda)$, $\beta(\lambda)$ with

$\alpha^2 + \beta^2 = 1$ so that

$$\|\Omega_0 - \alpha(\lambda)\zeta_a - \beta(\lambda)\zeta_b\|_2 \to 0.$$

Since $\Omega_1$ is orthogonal to $\Omega_0$, we have

$$\|\Omega_1 - \beta(\lambda)\zeta_a + \alpha(\lambda)\zeta_b\|_2 \to 0.$$

Hypothesis (4) implies that $\alpha(\lambda)\beta(\lambda)$ is bounded below as $\lambda \to \infty$, so that we can obtain upper and lower bounds on $\alpha, \beta$. Let

$$\varphi_\lambda(x) = \lambda^{-\nu/2}\alpha(\lambda)^{-1}\Omega_0\big(\lambda^{-\frac{1}{2}}(x+a)\big),$$

$$\tilde{\varphi}_\lambda(x) = \lambda^{-\nu/2}\beta(\lambda)^{-1}\Omega_1\big(\lambda^{-\frac{1}{2}}(x+a)\big).$$

Then both $\varphi_\lambda$ and $\tilde{\varphi}_\lambda$ approach $\kappa_a$ in $L^2_{\text{loc}}$ as $\lambda \to \infty$ and they solve suitable Schrodinger equations for which we can use the last lemma; so $\varphi_\lambda \to \kappa_a$ uniformly on compacts and so $|g_\lambda(x) - \alpha(\lambda)/\beta(\lambda)| \to 0$ uniformly on $\{x|\ |x-a| \le \lambda^{-1/2}\}$ and similarly $|g_\lambda(x) + \alpha(\lambda)/\beta(\lambda)| \to 0$ uniformly on $\{x|\ |x-b| \le \lambda^{-1/2}\}$. Since $\alpha, \beta$ are uniformly bounded away from 0 and $\infty$, the desired result holds. □

We are now able to complete the proof of Theorem 1.5:

*Proof of Theorem* 1.5 (*lower bound*) (given Thm. 2.3). Let $\gamma(t)$ be the geodesic from $a$ and $b$. Then, by definition of $\rho$:

$$\max\{\min(\rho(\gamma(t),a), \rho(\gamma(t),b))\} = \tfrac{1}{2}\rho(a,b)$$

(and occurs at the point where $\gamma$ intersects the geodesic bisector of $a, b$). In particular, for any $\varepsilon$, we can find $\delta$ small and a smooth non-intersecting curve $\tilde{\gamma}$ from $a$ to $b$ so that

$$\max\{\min(\rho(x,a), \rho(x,b)) | \operatorname{dist}(x,\tilde{\gamma}) \le \delta\} \le \tfrac{1}{2}\rho(a,b) + \varepsilon$$

and so, by Theorem 2.3, we can be sure that for all $\lambda$ large we have

$$|\Omega_0(x)|^2 \ge e^{-\lambda[\rho(a,b)+3\varepsilon]} \tag{2.8}$$

for all $x \in T_\varepsilon$ where

$$T_\varepsilon = \{x | \operatorname{dist}(x,\tilde{\gamma}) \le \delta\}.$$

By shrinking $T_\varepsilon$ slightly if necessary, we can find smooth coordinates $y = (y_1, y_\perp)$ on a neighborhood of $T_\varepsilon$, so that $T_\varepsilon = \{y|\ |y_\perp| \le 1\}$, $\gamma \subset \{y | y_\perp = 0\}$ and $a = (0,0)$, $b = (1,0)$. Since these coordinates are smooth and $\gamma$ runs from $a$ to $b$, we can find $C$ so that for $\lambda$ large,

$$\big\{y | y_1 = 0, |y| \le C\lambda^{-\frac{1}{2}}\big\} \subset \big\{x|\ |x-a| \le \lambda^{-\frac{1}{2}}\big\},$$

$$\big\{y | y_1 = 1, |y| \le C\lambda^{-\frac{1}{2}}\big\} \subset \big\{x|\ |x-b| \le \lambda^{-\frac{1}{2}}\big\}.$$

Let $T^{(\lambda)} = \{y \mid |y_\perp| \le C\lambda^{-1/2}\}$. Then, by (2.4) and the definition of $g_\lambda$,

$$E_1(\lambda) - E_0(\lambda) = \tfrac{1}{2}\int |\nabla g_\lambda|^2 |\Omega_0(\lambda, x)|^2\, d^\nu x \tag{2.9}$$

$$\ge \tfrac{1}{2}\int_{T^{(\lambda)}} |\nabla g_\lambda|^2 |\Omega_0|^2\, d^\nu x.$$

Now for $\lambda$ large, we can use the lower bound (2.8) on $\Omega_0$, and change coordinates from $x$ to $y$, bounding a Jacobian from below by a constant $J$. Moreover, we can bound $\nabla g_\lambda \equiv \partial g_\lambda/\partial x$ by $\partial g_\lambda/\partial y$ since $|\partial g_\lambda/\partial y| \le \|\partial x/\partial y\|\, |\partial g_\lambda/\partial x|$ and $\|\partial x/\partial y\|^{-1}$ is bounded by a constant $H$. The result is

$$E_1(\lambda) - E_0(\lambda) \ge \tfrac{1}{2} JH e^{-\lambda(\rho(a,b)+3\varepsilon)} \int_{|y_\perp| \le C\lambda^{-1/2}} dy_\perp \int_0^1 dy_1 |\partial g_\lambda/\partial y|^2. \tag{2.10}$$

Next, we note that

$$|g_\lambda(0, y_\perp) - g_\lambda(1, y_\perp)|^2 = \left|\int_0^1 dy_1 (\partial g_\lambda/\partial y_1)(y_1, y_\perp)\right|^2 \tag{2.11}$$

$$\le \int_0^1 dy_1 |\partial g_\lambda/\partial y|^2$$

by the Schwarz inequality. By Lemma 2.5, the left side of (2.11) is bounded from below by $(4C)^2$ for a suitable constant, so that (2.10) yields

$$|E_1(\lambda) - E_0(\lambda)| \ge (\text{const})\lambda^{-(\nu-1)/2} e^{-\lambda(\rho(a,b)+3\varepsilon)};$$

so $\underline{\lim}\, -(1/\lambda)\ln|E_1 - E_0| \ge -(\rho(a,b) + 3\varepsilon)$ yielding the required lower bound. □

If one looks at our proofs of the upper and lower bounds, one sees that the right side of (2.9) is dominated by the contribution of the integral to a very small neighborhood of the point where the geodesic crosses the geodesic bisector of $(a, b)$. Presumably, higher order asymptotics on $E_1 - E_0$ will require assumptions on the number of geodesics from $a$ to $b$ and an analysis near these special points where $\rho(x, a) = \rho(x, b) = \frac{1}{2}\rho(a, b)$.

## 3. Large deviations and semigroup bounds on eigenfunctions

In this section, we prove Theorems 2.1 and 2.3 and thereby complete the proofs of Theorems 1.3 and 1.5. We will use Brownian motion methods; in the next section we give PDE proofs. We begin with Theorem 2.1 which is elementary. Since Brownian motion most naturally controls the kernel of the semigroup $e^{-tH}$, we must begin with a bound on eigenfunctions in terms of that kernel.

PROPOSITION 3.1. *Let $H = -\frac{1}{2}\Delta + V$ with $V$ bounded below and continuous. Let $H\psi = E_0\psi$ with $\|\psi\|_{L^2} = 1$. Then, for any $s$:*

$$|\psi(x)|^2 \leq e^{sE_0}(e^{-sH})(x, x). \tag{3.1}$$

*Proof.* In [49], a "local spectral density," $dL(x, E)$, is introduced, so that
(i) $dL$ is a positive measure,
(ii) $dL(x, \{E_1\}) = \sum |\psi_j(x)|^2 \, dx$,
the sum being over a basis of $L^2$ eigenfunctions for $H\psi = E_1\psi$. $E_1$ is any fixed $E$.
(iii) $\int f(x)g(E)\, dL(x, E) = \mathrm{Tr}(f(x)g(H))$ for all continuous functions $f, g$ vanishing sufficiently rapidly at infinity.

In particular, choosing $g(E) = e^{-sE}$, $f$ a positive continuous function and using the fact that $e^{-sH}(x, y)$ is continuous (see [51]), we have

$$\begin{aligned}
\int f(x)|\psi(x)|^2\, dx &\leq \int f(x)\, dL(x, \{E_0\}) && \text{(by (ii))}\\
&\leq \int f(x) e^{sE_0} e^{-sE}\, dL(x, E) && \text{(by (i))}\\
&= \mathrm{Tr}\big(f(x)e^{-sH}\big)e^{sE_0} && \text{(by (iii))}\\
&= \int f(x) e^{-sH}(x, x) e^{sE_0}.
\end{aligned}$$

Since $e^{-sH}(x, x)$ and $|\psi(x)|^2$ are continuous [49], (3.1) follows. □

*Remarks* 1. Since the above proof is somewhat abstract, it is worth giving the simpler proof for the case where $(H + i)^{-1}$ is compact (as happens if $V(x) \to \infty$ at infinity which is usual for many cases of interest in Thms. 1.3, 1.5). In that case, $H$ has a complete set of eigenfunctions $\psi_n(x)$ and

$$e^{-tH}(x, y) = \sum_{n=1}^{\infty} e^{-tE_n}\psi_n(x)\psi_n(y).$$

If $E_0 = E_{n_0}$ and $\psi = \psi_{n_0}$ then (3.1) just makes the obvious assertion that

$$|\psi_{n_0}(x)|^2 e^{-sE_{n_0}} \leq \sum_n e^{-sE_n}|\psi_n(x)|^2.$$

From this special case and a limiting argument, one can obtain the result if $E_0 < \inf(\text{ess spec}(H))$.

2. In the notation of [49], only $V_- \in K_\nu$, $V_+ \in K_\nu^{\text{loc}}$ are needed.

The other result that we will need is the Feynman-Kac formula for $e^{-tH}(x, y)$. For a proof and background, see [50]. Let $E_{x,y;t}(\cdot)$ denote expectation value for the process $\{b(s)\}_{0 \leq s \leq t}$ where $\gamma(s) = b(s) - t^{-1}sy - (1 - t^{-1}s)x$ are

Gaussian random variables of mean 0 and covariance

$$E_{x,y;t}(\gamma(s_1)\gamma(s_2)) = s_1(1 - t^{-1}s_2) \quad \text{if } 0 \le s_1 \le s_2 \le t$$

(Brownian motion starting at $x$ and conditioned to have $b(t) = y$). Let $e^{-tH_0}(x, y) \equiv (2\pi t)^{-\nu/2}\exp(-(x-y)^2/2t)$ be the integral kernel of $e^{t\Delta/2}$ and let $H = -\frac{1}{2}\Delta + V$ (with say $V \ge 0$ and continuous). Then the Feynman-Kac formula says that

$$e^{-tH}(x, y) = e^{-tH_0}(x, y)E_{x,y;t}\left(e^{-\int_0^t V(b(s))\,ds}\right). \tag{3.2}$$

Later we will use a measure $\tilde{E}_{x,y;t}$ defined on functions from $[0, 1]$ to $R^\nu$ by letting $\tilde{b}(s) = b(st)$, $0 \le s \le 1$, and using the measure $E_{x,y;t}$. By scaling, the process for $\tilde{E}_{0,0;t}$ is just the process for $\tilde{E}_{0,0;1}$ multiplied by $\sqrt{t}$.

The proof of Theorem 2.1 requires an elementary fact about Brownian paths; let $P_{x,y;t}$ denote probabilities relative to the above process. Then

PROPOSITION 3.2.

$$P_{x,x;t}\left(\sup_{0\le s\le t} |b(s) - x| \ge R\right) \le C_1\exp(-R^2/C_2 t) \tag{3.3}$$

*for some* $C_1, C_2$.

*Proof.* Without loss of generality, one can take $x = 0$. If $\sup_{0\le s\le t}|b(s)| \ge R$, then for some choice of $i = 1, 2, \ldots, \nu$ and $\pm$, we must have that $\sup_{0\le s\le t} \pm b_i(s) \ge R/\sqrt{\nu}$, so that

$$P_{0,0;t}\left(\sup_{0\le s\le t} |b(s)| \ge R\right) \le 2\nu P_{0,0;t}\left(\sup_{0\le s\le t} b_1(s) \ge R/\sqrt{\nu}\right).$$

But, by the connection of restricted Brownian motion to Dirichlet Hamiltonians and the method of images [50], one has

$$P_{0,0;t}\left(\sup_{0\le s\le t} b_1(s) \ge a\right) = e^{-2a^2/t}. \qquad \square$$

With these preliminaries, we have

*Proof of Theorem* 2.1. Fix $x, \lambda, t$ and $R_0$. Then, by (3.1) and (3.2)

$$\begin{aligned}|\Omega_0(\lambda; x)|^2 &\le e^{tE_0(\lambda)}(2\pi t)^{-\nu/2}E_{x,x;t}\left(e^{-\lambda^2\int_0^t V(b(s))\,ds}\right)\\ &\le e^{tE_0(\lambda)}(2\pi t)^{-\nu/2}\Big[P_{x,x;t}\left(\sup_{0\le s\le t}|b(s) - x| \ge R\right)\\ &\qquad + \exp\left(-\lambda^2 t \inf\{V(y) \mid |y - x| \le R\}\right)\Big],\end{aligned}$$

where the first term comes from those paths which go a distance at least $R$ from

$x$ (and we use $V \geq 0$) and the second from paths that stay near $x$. Choosing $t = T/\lambda$ and using the last proposition, we obtain

$$|\Omega_0(\lambda; x)|^2 \leq \exp(TE_0(\lambda)/\lambda)(2\pi T/\lambda)^{-\nu/2}$$
$$\times\left[C_1\exp\left(-\lambda R^2/C_2 T\right) + \exp(-\lambda T m(x; R))\right]$$

where $m(x; R) = \inf\{V(y) \mid |y - x| \leq R\}$. Let $\delta > 0$ and $R_1$ be chosen so that $V(y) \geq \delta$ if $|y| > R_1$ and let $R_0 = 2R_1$. If $|x| \geq R_0$, we choose $R = \frac{1}{2}|x|$ (so that $m(x, R) \geq \delta$) and $T = R$ and find (2.3) using the fact that $E_0(\lambda)/\lambda$ is bounded as $\lambda \to \infty$. To get (2.2), choose $R = \varepsilon/2$ (so if $|x - a| \geq \varepsilon$ and $|x - b| \geq \varepsilon$, then $m(x, R) \geq \delta$ for some $\delta > 0$ independent of which $x$) and $T = 1$. □

The proof of Theorem 2.3 is more subtle, requiring "the method of large deviations." For a detailed presentation of these ideas and their development, see the notes of Varadhan [56]. The basic idea in our context is to note that as $t \downarrow 0$, the measures $\tilde{E}_{x,y;t}$ on function space approach a point mass at the line from $x$ to $y$. If $A$ is a set disjoint from this line, $\tilde{P}_{x,y;t}(A)$ should go to zero exponentially in $t^{-1}$ and modulo technicalities (upper bounds hold for open sets, lower bounds for closed sets and the continuity of $V$ handles that below),

$$\tilde{P}_{x,y;t}(A) \sim \exp\left(-t^{-1}\inf_{f\in A}\left[\frac{1}{2}\int_0^1 \mathring{f}(s)^2\,ds - \tfrac{1}{2}(x-y)^2\right]\right).$$

From this one obtains, using (3.2):

THEOREM 3.3. *Let $V$ obey hypotheses* (1)–(3). *Then for each fixed* $x, y; T$:

$$\lim_{\lambda\to\infty}\frac{1}{\lambda}\ln[\exp(-TH(\lambda)/\lambda)(x,y)] = -\alpha(x,y;T) \tag{3.4}$$

*where*

$$\alpha(x,y;T) = \min\left\{\frac{1}{2}\int_0^T \mathring{\gamma}^2(s)\,ds + \int_0^T V(\gamma(s))\,ds \,\middle|\, \gamma(0) = x, \gamma(T) = y\right\} \tag{3.5}$$

*and the limit in* (3.4) *is uniform in* $x, y, T$ *as they run through compact subsets of* $R^\nu \times R^\nu \times (0, \infty)$.

This is a direct consequence of Theorem 5.1 in Varadhan [56]; actually, it only needs the old results of Pincus [38] and Schilder [43]. To aid the reader, we describe a simple formal argument which suggests why (3.4) is true. Formally, one can think of $G_0(x, y; t)E_{x,y;t}(\cdot)$ as

$$\text{“}\textstyle\int\text{”}\, d^\infty b \exp\left(-\int_0^t \tfrac{1}{2}\mathring{b}(s)\,ds\right)\delta(b(0) - x)\delta(b(t) - y).$$

In terms of this formal notation,

$$\exp(-H(\lambda)t)(x,y) = \text{``}\!\int\!\text{''}\, d^\infty b\,\delta(b(0)-x)\delta(b(t)-y)$$

$$\times \exp\left(-\frac{1}{2}\int_0^t \mathring{b}(s)^2\,ds - \lambda^2\int_0^t V(b(s))\,ds\right).$$

Take $t = T/\lambda$ and $\gamma(s) = b(s/\lambda)$. Then a change of variables, when $\mathring{\gamma} = \lambda^{-1}\mathring{b}$, yields

$$\exp(-H(\lambda)T/\lambda)(x,y)) = \text{``}\!\int\!\text{''}\, d^\infty\gamma\,\delta(\gamma(0)-x)\delta(\gamma(T)-y)$$

$$\times \exp\left(-\lambda\left[\frac{1}{2}\int_0^T \mathring{\gamma}^2(u)\,du + \int_0^T V(\gamma(u))\,du\right]\right)$$

which formally suggests (3.4), (3.5).

To get the upper bound in Theorem 2.3, we need

LEMMA 3.4. $\lim_{T\to\infty}\alpha(x,x;T) = 2\min(\rho(x,a),\rho(x,b))$ *and the limit is uniform as $x$ runs through compact sets.*

*Proof.* The main idea is that as $T \to \infty$, the minimizing path for (3.5) must spend most of its time at $a$ or $b$ to avoid $\int_0^T V(\gamma(\cdot))\,ds$ becoming too large. For $T$ large, the minimizing path is essentially a minimum action path from $x$ to $a$ (or $b$) run for time $T/2$ and then its reverse. The details are straightforward and not even too tedious. □

*Proof of Theorem* 2.3 (*upper bound*). Given $\varepsilon$ and a compact $K \subset R^\nu$, first pick $T$ so that $\alpha(x,x;T) \geq 2\min(\rho(x,a),\rho(x,b)) - \varepsilon$ for all $x \in K$ and then $\Lambda$ and $\tilde{C}_\varepsilon$ so that $\exp(-TH(\lambda)/\lambda)(x,x) \leq \tilde{C}_\varepsilon \exp(-\lambda[\alpha(x,x;T) - \varepsilon])$ if $\lambda > \Lambda$. Let $C_\varepsilon = [\tilde{C}_\varepsilon \sup_{\lambda\geq\Lambda}\exp(TE_0(\lambda)/\lambda)]^{1/2}$. Thus, by Proposition 3.1, Theorem 3.3 and Lemma 3.4:

$$|\Omega_0(x)| \leq C_\varepsilon \exp[-\lambda(\min(\rho(x,a),\rho(x,b)) - \varepsilon)]$$

as required. □

As a preliminary to a proof of the lower bound, we need

PROPOSITION 3.5. *Under hypotheses* (1)–(4),

$$\Omega_0(x;\lambda) \geq C\lambda^{-\nu/2}\inf\left(e^{-sH(\lambda)}(y,x)\,\middle|\, |y-a| \leq \lambda^{-\frac{1}{2}}\right) \tag{3.6}$$

*for all $\lambda$ large. The estimate is uniform in $x$ and $s$.*

*Proof.* The proof of Lemma 2.5 shows that in the region $|y - a| \le \lambda^{-1/2}$, we have $\Omega_0(y; \lambda) \ge \tilde{C}\lambda^{\nu/2}$. Since $\Omega_0 \ge 0$ and

$$\Omega_0(x; \lambda) = e^{sE_0(\lambda)} \int e^{-sH(\lambda)}(x, y)\Omega_0(y; \lambda)\, dy$$

we obtain (3.6) by using $E_0 > 0$, and taking the contribution to the integral from the region $|y - a| \le \lambda^{-1/2}$. □

*Proof of Theorem* 2.3 (*lower bound*). By symmetry in $a, b$ and a compactness argument, it suffices to find for each $z$ and $\varepsilon$, a neighborhood $N$ of $z$ and $D_\varepsilon$ so that for $x \in N$,

$$\Omega_0(x; \lambda) \ge D_\varepsilon \lambda^{-\nu/2} \exp(-\lambda[\rho(x, a) + \varepsilon]). \tag{3.7}$$

First find $s$, so that $\alpha(z, a; s) \le \rho(z, a) + \frac{1}{3}\varepsilon$. This is possible since $\rho(x, y) = \inf_s \alpha(x, y; s)$. Since $\rho$ and $\alpha$ are continuous in their arguments, we can find $N$, a neighborhood of $z$, and $\delta$, so that for $x \in N$, and $|y - a| < \delta$:

$$\alpha(x, y; s) \le \rho(x, a) + \tfrac{1}{2}\varepsilon. \tag{3.8}$$

By Theorem 3.3, for some $\Lambda_0$ and $\tilde{D}_\varepsilon$:

$$\exp(-sH(\lambda)/\lambda)(x, y) \ge \tilde{D}_\varepsilon \exp(-[\alpha(x, y; s) - \tfrac{1}{2}\varepsilon]\lambda) \tag{3.9}$$

so long as $\lambda > \Omega_0$, $x \in N$ and $|y - a| < \delta$. Now (3.6), (3.8) and (3.9) yield the desired bound (3.7). □

## 4. Bounds on the eigenfunctions by Agmon's methods

In this section we give an alternate proof of Theorems 2.1 and 2.3. We do this partly for the benefit of readers unfamiliar with Brownian motion, since the proof here uses only "standard methods" of PDE's and partly because, in some situations (see Section 7), this method may be easily applicable while the probabilistic method requires added work.

Our methods closely parallel those that Agmon used to control the decay of Schrödinger $N$-body wave functions at large $|x|$; the upper bounds follow [1], [2]; the lower bounds follow some as yet unpublished work [3]. One can also get the lower bound à la Carmona-Simon [13] using only Jensen's inequality in function space (which is also how one gets the lower bound in Thm. 3.3).

We illustrate the upper bound methods by proving that part of Theorem 2.1 which is not contained in Theorem 2.3:

THEOREM 4.1. *Under hypotheses* (1)–(3), *there exist* $R_0$, $C$ *and* $d > 0$ *so that for* $\lambda$ *large and* $|x| > R_0$,

$$|\Omega_0(\lambda; x)| \le Ce^{-d|x|\lambda}.$$

*Proof.* Pick $R_0$ and $\delta$ so that $V(x) \geq \delta^2$ if $|x| > \frac{1}{4}R_0$. Let $\varphi$ be a function obeying for some $R_1$:

(a) $\varphi$ *is* $L^\infty$;

(b) $\varphi(x) = x$ *if* $|x| \leq R_1$, $0 \leq \varphi'(s) \leq 1$;

(c) $\varphi'(x) = 0$ *if* $|x| \geq 2R_1$;

and let $\rho(x) = \delta\varphi(|x|)$ so that $|\nabla\rho|^2 \leq \delta^2$ and $\rho$ is bounded and $C^\infty$ away from $|x| = 0$. Let $\psi$ be real-valued and supported away from $|x| = 0$. Then

(4.1)

$$\begin{aligned}\left(e^{\lambda\rho}\psi, (H(\lambda) - E_0(\lambda))e^{-\lambda\rho}\psi\right) &= \left(\psi, \left[-\tfrac{1}{2}(\nabla - \lambda\nabla\rho)^2 + \lambda^2 V - E_0(\lambda)\right]\psi\right)\\ &\geq \left(\psi, \left[\lambda^2\left(V - \tfrac{1}{2}(\nabla\rho)^2\right) - E_0(\lambda)\right]\psi\right) \geq \|\psi\|^2,\end{aligned}$$

so long as $\psi$ is supported in the region $\{x \mid |x| > \frac{1}{4}R_0\}$ and so long as $\lambda$ is so large that $\lambda^2(\frac{1}{2}\delta^2) - E_0(\lambda) \geq 1$. In the first inequality above one uses $-\Delta \geq 0$ and the fact that $(\psi, [\nabla(\nabla\rho) + (\nabla\rho)\nabla]\psi) = 0$ since it is both real and imaginary.

Now let $\eta$ be a function with $1 - \eta \in C_0^\infty$, $\eta = 0$ if $|x| < \frac{1}{4}R_0$, $\eta = 1$ if $|x| > \frac{1}{2}R_0$ and choose $\psi = e^{\lambda\rho}\eta\Omega_0$. Then the left side of (4.1) becomes:

$$\left(e^{2\lambda\rho}\eta\Omega_0, -(\nabla\eta)(\nabla\Omega_0) - \tfrac{1}{2}(\Delta\eta)\Omega_0\right).$$

This quantity is bounded independently of $R_1$ so we can take $R_1 \to \infty$ and replace $\rho$ by $\delta|x|$. Moreover, since $\|\Omega_0\|_2 = 1$ and $\|\nabla\Omega_0\|_2$ only grows linearly in $\sqrt{\lambda}$, (4.1) implies

$$\int_{|x| > \frac{1}{2}R_0} e^{2\delta\lambda|x|}|\Omega_0|^2 \leq Ce^{\delta\lambda R_0}\lambda$$

for $\lambda$ large. Thus

$$\int_{|x| > R_0} e^{\delta\lambda|x|}|\Omega_0(\lambda, x)|^2\, d^\nu x \leq C\lambda.$$

Since $\Omega_0$ is subharmonic in $|x| > R_0$, we can bound $\Omega_0$ by its average over a unit ball and find that for $\lambda$ large

$$|\Omega_0(\lambda, x)| \leq Ce^{-\frac{1}{3}\delta|x|\lambda}$$

as required. □

The same idea yields:

*Second proof of Theorem* 2.3 (*upper bound*). We sketch the idea, since the technical details are so close to those above. Let

$$\tilde{\varphi}(x) = \min(\rho(x,a), \rho(x,b)).$$

Then

$$|\tilde{\varphi}(x) - \tilde{\varphi}(y)| \le \rho(x,y) \le \int_0^1 d\theta \sqrt{2V(\theta x + (1-\theta)y)},$$

so that convolutions of $\tilde{\varphi}$ will have gradients very close to being bounded by $\sqrt{2V(x)}$. As a result, given $\delta$ and $R$, one can find $\varepsilon$ and $\varphi$ (by convolution and cutoff from $\tilde{\varphi}$) so that for $|x| < R$

$$(1-\delta)\tilde{\varphi}(x) \le \varphi(x) \le (1+\delta)\tilde{\varphi}(x), \tag{4.2}$$

$\varphi$ is bounded and

$$|(\nabla\varphi)(x)| \le (1-\varepsilon)\sqrt{2V(x)} \tag{4.3}$$

for all $x$. As in the last proof, one finds that using (4.3):

$$\left(e^{\lambda\varphi}\psi, (H(\lambda) - E(\lambda))e^{-\lambda\varphi}\psi\right) \ge \left(\psi, \lambda^2 \varepsilon V(x)\psi\right);$$

and so for any $\kappa > 0$, we can find $\Lambda > 0$, so that for $\lambda > \Lambda$

$$\|\psi\|^2 \le \left(e^{\lambda\varphi}\psi, (H - E)e^{-\lambda\varphi}\psi\right) \tag{4.4}$$

so long as $\psi(x) = 0$ if $|x - a| < \kappa$ or $|x - b| < \kappa$. Given $\kappa$, we can find $\kappa_0$ much smaller so that $|x - a| \ge \kappa$, $|x - b| \ge \kappa$, and $|y - a| \le \kappa_0$ or $|y - b| \le \kappa_0$ implies $\varphi(y) < \delta\varphi(x)$. As in the last proof, (4.4) then yields

$$\begin{aligned}|\Omega_0(\lambda, x)| &\le C\exp(-(1-\delta)\varphi(x))\\ &\le C\exp\left(-(1-\delta)^2\tilde{\varphi}(x)\right),\end{aligned}$$

so long as $|x| < R$, $|x - a| > \kappa$, $|x - b| > \kappa$. This is the desired bound. □

Agmon's lower bound technique [3] relies on an elementary lemma (Lemma 4.3 below) which in turn relies on a simple application of the maximum principle.

LEMMA 4.2 (e.g. Simon [45]). *Let $\psi(x) \ge 0$, $\Omega(x) \ge 0$ on a neighborhood of a region $D$ so that on this neighborhood, $\psi, \Omega$ are continuous, obeying*

$$\tfrac{1}{2}\Delta\Omega(x) = W(x)\Omega(x); \quad \tfrac{1}{2}\Delta\psi(x) = U(x)\psi(x)$$

*in distributional sense and $0 \le W(x) \le U(x)$. If $\Omega \ge \psi$ on $\partial D$, then $\Omega \ge \psi$ on all of $D$.*

*Proof.* Let $\eta = \Omega - \psi$ and $D_0 = \{x | \eta(x) < 0\}$. Notice that since $\eta \ge 0$ on $\partial D$, we have that $\eta \equiv 0$ on $\partial D_0$. Now, on $D_0$:

$$\tfrac{1}{2}\Delta\eta = W\Omega - U\psi = W\eta + (W - U)\psi \le 0$$

since $\eta \le 0$ on $D_0$ and $W - U \le 0$. Thus $\eta$ is superharmonic on $D_0$ and in particular, $\min_{x \in D_0}\eta(x)$ occurs on $\partial D_0$, so that $\eta \ge 0$ on $D_0$ and $D_0$ is empty. □

The following is basic to Agmon's lower bound technique [3]:

LEMMA 4.3. *Let $\eta$ be the lowest eigenfunction of $-\frac{1}{2}\Delta$ with Dirichlet boundary conditions on the unit ball in $\nu - 1$ dimensions, normalized so that $\|\eta\|_\infty = 1$, and let $e_0$ be the corresponding eigenvalue. Let $d = \min_{|y|\le 1/2}\eta(y) > 0$. Let $D_0$ be the cylinder in $R^\nu$*

$$D_0 = \{x = (x_1, x_\perp)|0 \le x_1 \le a(1+\delta);\quad |x_\perp| \le R\}.$$

*Let $\Omega$ obey $\frac{1}{2}\Delta\Omega = W(x)\Omega$ on $D_0$ with $W \ge 0$ and $\Omega \ge 0$ there. Define $\alpha$ by*

$$\frac{1}{2}\alpha^2 = \sup_{x\in D_0}\left[\frac{e_0}{R^2} + W(x)\right]. \tag{4.5}$$

*Then*

$$\min\{\Omega(x)|x_1 = a, |x_\perp| \le \tfrac{1}{2}R\}$$
$$\ge de^{-\alpha a}(1 - e^{-2\delta\alpha a})\min\{\Omega(x)|x_1 = 0, |x_\perp| \le R\}.$$

*Proof.* Without loss of generality, we can normalize $\Omega$ so that

$$\min\{\Omega(x)|x_1 = 0, |x_\perp| \le R\} = 1. \tag{4.6}$$

Let

$$\psi(x) = \eta(x_\perp/R)\left[e^{-\alpha x_1} - e^{-2\alpha a(1+\delta)}e^{\alpha x_1}\right].$$

Then $\psi$ vanishes on the sides of the cylinder and on the face $x_1 = a(1+\delta)$; so by (4.6), $\psi \le \Omega$ on $\partial D$. Moreover

$$\frac{1}{2}\Delta\psi = \left[-\frac{E_0}{R^2} + \frac{1}{2}\alpha^2\right]\psi$$

and, by (4.5), $W \le [-(E_0/R^2) + \frac{1}{2}\alpha^2]$ on $D_0$. Thus Lemma 4.3 is applicable and yields the desired result. □

*Second proof of Theorem* 2.3 (*lower bound*). We prove the necessary bound at a fixed $x$; the uniformity for points near $x$ is a simple consequence of adding one more linear segment to the argument below. Given $\varepsilon$, find first a piecewise linear path, $\gamma$, from $a$ to $x$ with $\int_0^1|\dot\gamma(s)|\sqrt{2V(\gamma(s))}\,ds \le \rho(x,a)(1+\varepsilon/3)$. Then by further subdividing the linear paths into small segments, one can find $x_0 = x, x_1, x_2, \ldots, x_n = a$, $\delta$ and $R_0$ small so that if $D_i$ is the cylinder of radius $R_0$ orthogonal to the segment from $x_{i-1}$ to $x_i$, with axis this segment increased in size by $(1+2\delta)$, and if $\nu_i = \sup_{x\in D_i}V(x)$, then

$$\sum\sqrt{2\nu_i}|x_i - x_{i-1}| \le \rho(x,a)\left(1 + \frac{2\varepsilon}{3}\right). \tag{4.7}$$

We already know that within a distance $C\lambda^{-1/2}$ of $a$, $\Omega_0(\lambda, x) \ge 1$ for $\lambda$ large (see the proof of Lemma 2.5). We can find points this far from $a$ outside the region where $\lambda^2 V(x) - E_0(\lambda) \leq 0$ (this region also shrinks as $\lambda^{-1/2}$) and then apply Lemma 4.3 in a sequence of cylinders. We start out with $R_1 = \frac{1}{2}\lambda^{-1/2}$ so that we know $\Omega_0(\lambda, x) \geq 1$ on the first cylinder's base. Then $R_j = 2^{-j}\lambda^{-1/2}$ and so $R_j \geq 2^{-n}\lambda^{-1/2}$ for all $j$ and $R_j \leq R_0$ for all $j$ if $\lambda$ is sufficiently large. If $\alpha_j(\lambda)$ is defined by

(4.8) $$\tfrac{1}{2}\alpha_j(\lambda)^2 = 2^n e_0 \lambda + \nu_j \lambda^2,$$

then Lemma 4.3 yields:

$$\Omega_0(\lambda, x) \geq e^{-\Sigma \alpha_i |x_i - x_{i-i}|} d^n \left( \prod_{i=1}^{n} \left[1 - e^{-2\delta\alpha_i(\lambda)}\right] \right).$$

Choose $\Lambda$ so large that

$$\exp\left(-2\delta\sqrt{2^{n+1}\Lambda e_0}\right) \leq \tfrac{1}{2}.$$

Then for $\lambda \geq \Lambda$ and large enough for $R_j \leq R$ and $\Omega_0 \geq 1$ on the first cylinder base, we have that

$$\Omega_0(\lambda, x) \geq e^{-\Sigma \alpha_i |x_i - x_{i-1}|} \left(\tfrac{1}{2} d\right)^n.$$

By (4.7) and (4.8), it is easy to see that for $\lambda$ large

$$\Omega_0(\lambda, x) \geq \exp(-\lambda\rho(x, a)(1 + \varepsilon)).$$

A similar argument with $b$ replacing $a$ yields the desired lower bound. □

## 5. Extensions (a): Manifolds

Molchanov [37] (following in part Varadhan [54], [55]) has discussed large deviations for Brownian motion on a complete Riemannian manifold, $M$. Using his results one can treat eigenvalue degeneracy for operators of the form:

(5.1) $$H(\lambda) = \tfrac{1}{2}L + \lambda^2 V.$$

Here $L$ is the Laplace Beltrami operator. For simplicity (see below), we suppose $M$ is compact. $V$ is then supposed $C^\infty$ with $V \geq 0$, $V(x) = 0$ only at $x = a, b$ and these minima are non-degenerate. The Agmon metric is then $\sqrt{2V(x)}\, g$ where g is the original metric on $M$. Note that $\rho(a, b)$ is the distance from $a$ to $b$ in the Agmon metric. Using the method of Sections 2, 3 with [37] replacing Schider [43], one obtains:

THEOREM 5.1. *Let $M$ be a compact Riemannian manifold and let $V$ obey the above hypotheses. Suppose that* $\underline{\lim}\|j_a\Omega_0\|\,\|j_b\Omega_0\| > 0$. *Then*

$$\lim_{\lambda\to\infty} -\frac{1}{\lambda}\ln|E_1 - E_0| = \rho(a, b),$$

*the distance in the Agmon metric ($\neq$ original metric).*

Compactness can be replaced by suitable restrictions on the behavior at infinity: One needs to know $\mathrm{vol}\{x \mid d(x,a) < R\}$ grows less than exponentially in $R$ and enough information to get the analog of Theorem 2.1.

## 6. Extensions (b): Complicated zeros, degenerate minima, excited states

In this section, we describe a number of extensions in a series of remarks. Except for (1), (3), we only pay attention to upper bounds on eigenvalue splittings; lower bounds may be false for general excited states, and in any event, seem to be very difficult to prove.

(1) $-\frac{1}{2}\Delta + \lambda^2 V + \lambda W$. Suppose that $W$ is a $C^\infty$ function which is bounded (actually, only $W(V+1)^{-1}$ need be bounded) and let

$$H(\lambda) = -\tfrac{1}{2}\Delta + \lambda^2 V + \lambda W. \tag{6.1}$$

In the asymptotics for $e^{-TH(\lambda)/\lambda}(x,y)$, it is easy to see that the $W$ term at most affects the path integral by a multiplicative constant between $e^{-T\|W\|_\infty}$ and $e^{T\|W\|_\infty}$, so that the leading asymptotics of $e^{-TH(\lambda)/\lambda}(x,y)$ are unchanged, i.e., are determined by the action with $V$ alone (and $W$ doesn't matter). From there the arguments of Sections 2, 3 imply that Theorem 1.5 extends with $\rho$ the $V$-Agmon metric.

This extension is important because there are many examples of the form (6.1) where hypothesis (4) holds (i.e., where the ground state asymptotically lives in both wells) even though $H(\lambda)$ has no symmetry: Let $V$ be any function obeying hypotheses (1)–(3) and $W$ any bounded $C^\infty$ function with $W(a) \neq W(b)$. We claim (following an idea of Davies [22]) that one can find $\alpha(\lambda)$ so that $\alpha(\lambda)/\lambda \to c_0$, a computable constant and so that the ground state $\Omega_0$ for

$$H(\lambda) = -\tfrac{1}{2}\Delta + \lambda^2 V + \alpha(\lambda) W \tag{6.2}$$

obeys hypothesis (4). For, by [46], if $\alpha(\lambda)/\lambda \to c$, the ground state energy, $E_0(\lambda)$ of (6.2) obeys

$$E_0(\lambda)/\lambda \to \min(e_a, e_b),$$

$$e_a = \mathrm{Tr}\left(\left[\partial^2 V/\partial x_i\, \partial x_j(a)\right]^{\frac{1}{2}}\right) + W(a)c \tag{6.3}$$

and similarly for $b$. In (6.3) we mean the trace of the square root of the positive matrix $\partial^2 V/\partial x_i\, \partial x_j(a)$. Moreover, by [46], if $e_a < e_b$, $\Omega_0$ is asymptotically concentrated in well $a$. If $c$ is adjusted to a value $c_0$ so that $e_a = e_b$, then for $\lambda$ large and $\alpha(\lambda)/\lambda$ near $c_0$, the ground state shifts from being concentrated in $a$ to being concentrated in $b$ as $\alpha(\lambda)$ varies; and since one can show it is uniformly concentrated in the union of wells, one sees by continuity that $\alpha(\lambda)$ can be adjusted so that $\alpha(\lambda)/\lambda \to c_0$, and so that hypothesis (4) holds. As explained by Davies [22], what happens here is that the asymptotic levels $e_a(c)$ and $e_b(c)$ cross and since the vacuum is non-degenerate, there must be an avoided crossing, but one that is exponentially close. The Agmon metric shows how close.

The next example of this genre shows the complications that can occur when there are multiple minima. One might think that if $V$ has multiple minima but that asymptotically $\Omega_0$ is concentrated only in the two wells, $a, b$, then the eigenvalue splitting is determined by $\rho(a, b)$. As our proof of the lower bound in Theorem 1.5 shows, this always provided a lower bound on the splitting (i.e., an upper bound on $\lim -(1/\lambda)\ln|E_1 - E_0|$) but it may *not* give the answer. We consider the following construction suggested by Witten [57]. Let $f$ be a $C^\infty$ function on $R^\nu$ so that (i) $|f(x)| \geq C\ln|x|$ for large $x$, (ii) $\Delta f$ is bounded, (iii) $f \geq 0$, (iv) $f$ has exactly two zeros at $x = a, b$ which are non-degenerate minima, (v) $|\nabla f|$ is bounded away from zero near infinity. Let

$$(6.4)\qquad H(\lambda) = -\tfrac{1}{2}\left[e^{\lambda f}\dot{\nabla} e^{-2\lambda f}\nabla e^{\lambda f}\right] = -\tfrac{1}{2}\Delta + \tfrac{1}{2}\lambda^2(\nabla f)^2 - \tfrac{1}{2}\lambda\,\Delta f.$$

Then $\Omega_0 = C(\lambda)e^{-\lambda f}$ with $C(\lambda)$ a normalizing constant. Of course, $V = \frac{1}{2}(\nabla f)^2$ has zeros at *all* critical points but $\Omega_0$ is only concentrated at those which are absolute minima (in accordance with Witten's work on operators of the form (6.4)). Since $\Omega_0$ is explicitly known, using the ideas in Section 2 one can often compute the eigenvalue-splitting asymptotics. If

$$\alpha = \min_{\gamma}\left\{\max(f(\gamma(s)))|\gamma(0) = a,\ \gamma(1) = b\right\},$$

$$\beta = \max_{S}\left\{\min_{x\in S}(f(x))|S \text{ is a smooth surface separating } a \text{ and } b\right\},$$

and if $\alpha = \beta$ (this may well be true always), then $\lim -(1/\lambda)\ln|E_1 - E_0| = 2\alpha$. In one dimension $\alpha = \beta$ always, and $2\alpha$ is exactly the Agmon distance from $a$ to $b$ if $f$ has a unique local maximum in $[a, b]$, but if $f$ has a (non-zero) local minimum in $(a, b)$, then $2\alpha$ ($\equiv 2\max_{a\leq x\leq b} f(x)$) is strictly smaller than the distance in the Agmon metric! This example shows that while

$$\lim_{T\to\infty}\frac{1}{2}\alpha(x, x; T) = \min(\rho(x, y)|V(y) = 0)$$

is always the lower bound on $\lim_{\lambda\to\infty}(1/\lambda)\ln\Omega_0(x, \lambda)$, it may not be optimal.

(2) *Degenerate minima.* For the upper bounds we only used $E(\lambda)/\lambda$ bounded above and this is true even if the minima are degenerate, in which case $E(\lambda)$ may go to zero faster than linearly. Under enough hypotheses on the form of $V$ at the degenerate minima, it should be possible to get lower bounds also.

(3) *Multiple wells.* If $V(x)$ obeys the basic hypotheses but has non-degenerate minima at points $a_1, \ldots, a_m$, and if $\Omega_0$ has components in at least two wells, say in wells $a_j$, $j \in S$, then our proof of the upper bound works to prove

$$\overline{\lim} \frac{1}{\lambda} \ln |E_1 - E_0| \leq -\alpha$$

where $\alpha = \max_{j \in S}(\min_{i \neq j} \rho(a_i, a_j))$. For if $1 \in S$ and $\alpha = \min_{i \neq 1} \rho(a_i, a_1)$, we go through the proof of the upper bound using an $f$ with $\nabla f$ concentrated near the geodesic sphere about $a_1$ of radius $\frac{1}{2}\alpha$. One difficulty in general for multiple wells, illustrated by the last example in (1) above, is that we only know:

$$\overline{\lim_{\lambda \to \infty}} \frac{1}{\lambda} \ln \Omega_0(\lambda, x) \leq -\min\left(\rho(x, a_i) | \text{all } i\right),$$

$$\underline{\lim_{\lambda \to \infty}} \frac{1}{\lambda} \ln \Omega_0(\lambda, x) \geq -\min\left(\rho(x, a_i) | i \in S\right).$$

Another is that the geometry may be complicated so that the geodesic bisector between two $a$'s goes near a third.

One multiple minima situation where we can completely analyze splitting to the ground state is where the minima are related by a symmetry. For simplicity, we suppose that the symmetry group is a cyclic group of order $n$; i.e., for a rotation, $R$, of order $n$, $V$ is left invariant and there are exactly $n$ minima $a_1, \ldots, a_n$ cyclically permuted by the rotation. Then $L^2(R^\nu)$ breaks up into $n$ invariant subspaces $\mathscr{H}_j$ $(j = 0, 1, \ldots, n-1)$ with $f(Rx) = e^{2\pi i j/n} f(x)$. Let $E_0^{(j)}(\lambda)$ denote the lowest eigenvalue of $H(\lambda) \upharpoonright \mathscr{H}_j$. Then, we claim that for $j \neq 0$,

$$\lim_{\lambda \to \infty} -\frac{1}{\lambda} |E_0^{(j)}(\lambda) - E_0^{(0)}(\lambda)| = \min_{i \neq 1} \rho(a_i, a_1). \tag{6.4$'$}$$

We do not make any statement about lower bounds on the splitting $E_0^{(j)} - E_0^{(k)}$; indeed, $E_0^{(j)} = E_0^{(n-j)}$, by reality of $H(\lambda)$. Note that (6.4)$'$ follows from our general methods: We get the lower bound picking an $f$ which is $e^{2\pi i j/n}$ near $R^j a_1$ and with $\nabla f$ concentrated near the geodesic spheres about the $a_i$ of radius $\frac{1}{2}\min \rho(a_i, a_j)$ and the lower bounds by following a geodesic between two points with $\rho(a_k, a_l) = \min \rho(a_i, a_j)$.

The above provides another warning about multiple minima. As an example, consider a potential $V$ on $R^2$ with minima at the four points $(\pm 1, \pm 1)$ and symmetric under rotations about 0 of angle $j\pi/2$ and reflections in the lines $x = \pm y$. The symmetry group $C_{4v}$ has six elements and then irreducible representations of dimensions 2, 1, 1. In terms of the above analysis, $\mathscr{H}_1$ and $\mathscr{H}_3$ are combined in one space, $\mathscr{H}_\pi$, since they are linked by the reflections. This space, $\mathscr{H}_\pi$, has a basis of functions which are even under one reflection and odd under the other, and so a basis of functions which vanish on either $x = y$ or $x = -y$. Because of these zeros, one might naively think that what is relevant for the splitting of $E_0^{(0)}$ and $E_0^{(\pi)}$ is then the distance in the Agmon metric from $(1, -1)$ to $(-1, 1)$, but in fact it is the distance from $(1, -1)$ to $(1, 1)$ that counts.

(4) *Manifolds of minima.* For upper bounds on the splitting, it does not even matter that the set of zeros of $V$ is a finite set. Our upper bound proof immediately implies:

THEOREM 6.1. *Let $V$ be a smooth function obeying hypotheses* (1) *and* (2). *Suppose that*

$$\{x | V(x) = 0\} = S_1 \cup S_2$$

*where $S_1$, $S_2$ are disjoint sets. Let $j_1$ be a smooth function which is 1 near $S_1$ and 0 near $S_2$, and let $j_2$ be 1 near $S_2$ and 0 near $S_1$. Suppose* $\underline{\lim}\|j_1\Omega_0\|\,\|j_2\Omega_0\| > 0$. *Then*

$$\overline{\lim_{\lambda\to\infty}}\ \frac{1}{\lambda}\ln|E_1(\lambda) - E_0(\lambda)| \leq -\min\big(\rho(x, y) | x \in S_1, y \in S_2\big).$$

(5) *Excited states.* For excited states, one can only hope to get upper bounds on gaps. It is certainly possible in higher dimensions to construct examples using the construction of Davies (i.e., with $\lambda^2 V$ replaced by $\lambda^2 V + \alpha(\lambda)W$) where two eigenvalues, one in each well, are degenerate at all $\lambda$ with one eigenvector living in each well so that a suitable combination lives in both wells. For the ground state this could not happen since level crossing is not allowed (the ground state is simple). For excited states, one can arrange things so that symmetry allows crossing; e.g. in two dimensions, one can have $V(x, -y) = V(x, y)$, two minima at $(\pm 1, 0)$ and so that a state odd under $y \to -y$ living in the well at $(-1, 0)$ is degenerate with a state even under $y \to -y$ in the well at $(1, 0)$.

One should expect upper bounds with $\rho(a, b)$, but since we no longer have the variational formula in the simple form, $(f\Omega_0, \Omega_0) = 0$ implies $(E_1 - E_0) \leq (f\Omega_0, (H - E_0)f\Omega_0)/(f\Omega_0, f\Omega_0)$; we have to use an alternate method which loses a factor of 2. That is, we are only able to prove:

THEOREM 6.2. *Let $V$ obey hypotheses* (1)–(3), *and let* $|V(x)| \leq Ce^{A|x|}$ *for some $C$, $A$. Let $\Omega_n(\lambda, x)$ be an excited state eigenvector with*

(4′) $$\underline{\lim}\|j_a\Omega_n\|\,\|j_b\Omega_n\| > 0.$$

*Then there is another eigenvalue $E'(\lambda)$ so that*

$$\overline{\lim}\,\frac{1}{\lambda}\ln|E'(\lambda) - E(\lambda)| \leq -\tfrac{1}{2}\rho(a, b).$$

*Note.* In the symmetric case, hypothesis (4′) is automatic for all $n$. Thus all the low lying (i.e., fixed $n$) states are nearly doubly degenerate.

*Proof.* We first claim it suffices to find $f$ so that

$$(f\Omega_n, \Omega_n) = 0 \quad \text{and } \|(H - E_n)f\Omega_n\| / \|f\Omega_n\| = O\left(e^{-\frac{1}{2}\lambda(1-\varepsilon)\rho(a,b)}\right).$$

For this first equation implies

$$\|(H - E_n)f\Omega_n\| \geq \inf\left(E_n, \sigma\left(H \restriction \{\Omega_n\}^{\perp}\right)\|f\Omega_n\|\right.$$

(this is essentially a version of "Temple's inequality" [35]). We can, as in the ground state case, find $f_\lambda$ with $\nabla f_\lambda$ $\lambda$ independent and concentrated near the geodesic bisector and $(f\Omega_n, \Omega_n) = 0$, $\|f\Omega_n\| = O(1)$. Now, by Leibniz rule:

$$\|(H - E_n)f\Omega_n\| \leq \tfrac{1}{2}\|(\Delta f)\Omega_n\| + \|(\nabla f) \cdot \nabla\Omega_n\|.$$

By the support properties of

$$\nabla f, \|(\Delta f)\Omega_n\| = O\left(\exp\left(-\tfrac{1}{2}\lambda(1-\varepsilon)\rho(a,b)\right)\right).$$

By (2.7),

$$\int |\nabla f|^2 |\nabla\Omega_n|^2 = \frac{1}{2}\int \Delta\left((\nabla f)^2\right)|\Omega_n|^2 - \int (\Delta f)^2 (\Delta\Omega_n)\Omega_n.$$

Since $\Delta\Omega_n = [\lambda^2 V(x) - E(\lambda)]\Omega_n$ and $V(x)$ does not grow too fast, we see that $\|(\nabla f)\nabla\Omega_n\|$ is also $O(\exp[-\frac{1}{2}\lambda(1-\varepsilon)\rho(a,b)])$. □

## 7. Extensions (c): Highly excited states

In this section we want to consider a *sequence* of eigenstates $\psi(\lambda_n; x)$ with

(7.1) $$H(\lambda)\psi(\lambda_n; x) = E(\lambda_n)\psi(\lambda_n; x)$$

By picking a sequence, we are not varying $E$ continuously, which tends to force $E$ to grow only linearly in $\lambda$ (and so be "low lying" on the $\lambda^2$ scale). Instead, we want to pick $E(\lambda_n)$ so that

(7.2) $$E(\lambda_n)/\lambda_n^2 \to e_0;\ \lambda_n \to \infty.$$

We shall control decay of the eigenfunctions $\psi_n(\lambda, x)$ under certain circumstances and if there is symmetry, we shall obtain small splittings.

Given a number $e_0$, we let

$$A(e_0) = \{x \mid V(x) < e_0\},$$

and for any $x$, we define

$$\rho_{e_0}(x) = \inf\left(\int_0^1 \sqrt{2[V(\gamma(s)) - e_0]}\, |\dot{\gamma}(s)|\, \big|\, \gamma(0) = x,\ \gamma(1) \in A(e_0)\right).$$

THEOREM 7.1. *Suppose that* (1), (2) *hold and that* $A(e_0)$ *is compact* (*i.e.*, $V(x) > e_0$ *near infinity*). *Let* (7.1), (7.2) *hold. Then*

(i) $|\psi(\lambda_n; x)| \leq C\exp(-\delta|x|\lambda_n)$ *for some* $\delta > 0$ *and all large* $x$.

(ii) *For any* $R$ *and* $\varepsilon$, *there is a* $C(R,\varepsilon)$ *so that for all* $n$ *and all* $x$ *with* $|\lambda| < R$,

$$|\psi(\lambda_n, x)| \leq C(R,\varepsilon)\exp\big(-(1-\varepsilon)\lambda_n \rho_{e_0}(x)\big).$$

Before sketching the proof, let us get an eigenvalue splitting result from this. For simplicity, we suppose

$$V(x_1, x_2, \dots) = V(-x_1, x_2, \dots), \tag{7.3}$$

and that $V$ has a barrier, i.e.,

$$b = \inf_{x_1 = 0} V(x) > 0.$$

If $e_0 < b$, the sets $A^{\pm}(e_0) = A(e_0) \cap \{\pm x_1 > 0\}$ are a finite distance from one another. Set

$$\begin{aligned} d(e_0) &= \inf\left(\int_0^1 \sqrt{2[V(\gamma(s)) - e_0}\,|\dot\gamma(s)|\;|\gamma(0) \in A^+,\ \gamma(1) \in A^-\right) \\ &= 2\inf\{\rho_{e_0}(x) | x_1 = 0\}. \end{aligned}$$

Then Theorem 7.1 and the argument in the last section (subsection (5)), prove:

THEOREM 7.2. *Let* $V$ *obey hypotheses* (1), (2) *and equation* (7.3). *Suppose that* $A(e_0)$ *is compact, and* $e_0 < b$. *Then there is a sequence of states* $\tilde\psi(\lambda_n, x)$ *of the opposite symmetry to* $\psi$, *so that for any* $\varepsilon > 0$ *there is a* $D(\varepsilon)$ *with*

$$|\tilde E(\lambda_n) - E(\lambda_n)| \leq D(\varepsilon)\exp\left[-\left(\tfrac{1}{2} - \varepsilon\right)d(e_0)\right]. \tag{7.4}$$

As remarked in the last section, we believe that the $\frac{1}{2}$ in (7.4) is an artifact of the proof and can be replaced by 1.

We conclude with:

*Proofs of Theorem* 7.1. One can easily apply Agmon's method, since a smoothed out (and cut off near $\infty$) $\rho$ will obey $\frac{1}{2}(\nabla\rho)^2 \leq (V - e_0)(1+\varepsilon)$. Alternatively, one can *probably* use path integrals as follows: (a) First, noting that $(H+1)^{-N}\psi \leq (H_0 + 1)^{-N}\psi$, one sees that $\|\psi\|_\infty$ grows at most as a power of $\lambda$. (b) Running Brownian motion from $x$ to the stopping time $\tau = \inf\{s | b(s) \in A(e_0)\}$ and this power bound, one finds

$$|\psi(x)| \leq C\lambda^m E\left(\exp\left(-\lambda^2\int_0^\tau (V(b(s)) - e_0)\,ds\right)\right).$$

Since the integrated $V(b(s)) - e_0$ is positive if $s < \tau$, we can replace $\tau$ by $\tau \wedge (T/\lambda)$ for any fixed $T/\lambda$ ($a \wedge b = \inf(a, b)$). Thus

$$|\psi(x)| \leq C\lambda^m \mathscr{F}_\lambda(x, T)$$

where

$$\mathscr{F}_\lambda(x, T) = E\left(\exp\left(-\lambda^2 \int_0^{\tau \wedge (T/\lambda)} (V(b(s)) - e_0)\, ds\right)\right).$$

It should be true (but I have not tried to write down a proof since Agmon's proof works), that large deviations imply that

$$\lim_{\lambda \to \infty} -\frac{1}{\lambda} \ln \mathscr{F}_\lambda(x, T) = \inf\left\{\frac{1}{2}\int_0^{T} \dot{b}(s)^2\, ds + \int_0^T [V(b(s)) - e_0]\, ds \,\middle|\, \gamma(0) = x,\ \gamma(s) \notin A(e_0) \text{ for all } V\right\}$$

It is easy to see that as $T \to \infty$, this last expression converges to $\rho_{e_0}(x)$. □

Tunneling results in one dimension "part way up the barrier" are discussed in Fröman [24].

California Institute of Technology, Pasadena

*Notes added in proof.*

(1) Subsequent to this work, considerable further progress has been made on the problems treated here by B. Helffer and J. Sjöstrand: Multiple wells in the semiclassical limit, Commun. in PDE, to appear, and in some additional works. In particular, they do not lose the factor of 2 as we did in Theorem 6.2. They have gone beyond leading order under suitable hypotheses, and they have a much greater understanding of multiple wells than we found and stated in Chap. 6.

(2) We have now written two additional papers in this series. Paper III (Width of the ground state band in strongly coupled solids; Ann. Phys., to appear) deals with the multi-dimensional analog of the problem studied in [29]. Paper IV (The flea on the elephant; to be submitted to Commun. Math. Phys.) deals with the multi-dimensional analog of the problem studied in [33].

## References

[1] S. Agmon, Spectral properties of Schrödinger operators and scattering theory, Ann. Scuola Norm Sup. Pisa II **2** (1979), 151–218.

[2] ______, *Lectures on Exponential Decay of Solutions of Second Order Elliptic Equations. Bounds on Eigenfunctions of N-Body Schrödinger Operators*, Mathematical Notes, Princeton Univ. Press, Princeton, N.J., 1982.

[3] ______, in prep.

[4] J. Aguilar and J. Combes, A class of analytic perturbations for one-body Schrödinger Hamiltonians, Comm. Math. Phys. **22** (1971), 269–279.

[5] S. Albeverio, R. Hoegh-Krohn and L. Streit, Energy forms, Hamiltonians and distorted Brownian paths, J. Math. Phys. 18 (1977), 907–917.

[6] ______, Regularization of Hamiltonians and processes, J. Math. Phys. **21** (1980), 1636–1642.

[7] M. Ashbaugh and E. Harrell, Perturbation theory for shape resonances and large barrier potentials, Comm. Math. Phys. **83** (1982), 151–170.

[8] M. Ashbaugh and C. Sundberg, An improved stability result for resonances, Trans. AMS, to appear.

[9] E. Balslev and J. M. Combes, Spectral properties of many-body Schrödinger operators with dilatation analytic interactions, Comm. Math. Phys. **22** (1971), 280–294.

[10] C. Bender and T. T. Wu, Large order behavior of perturbation theory, Phys. Rev. Lett. **16** (1971), 461–465; Anharmonic oscillator, II. A study of perturbation theory in large order, Phys. Rev. **D7** (1973), 1620–1636.

[11] S. Breen, Large order perturbation theory for the anharmonic oscillator, UBC preprint.

[12] ______, Leading large order asymptotics for $(\phi^4)_2$ perturbation theory, Comm. Math. Phys., to appear.

[13] R. Carmona and B. Simon, Pointwise bounds on eigenfunctions and wave packets in $N$-body quantum systems, V. Lower bounds and path integrals, Comm. Math. Phys. **80** (1981), 59–98.

[14] S. Coleman, The uses of instantons, Proc. Internat. School of Physics, Erice, 1977.

[15] J. Combes, Asymptotic expansions for quantum mechanical bound state energies near the classical limit, pp. 22–38 in *Spectral and Scattering Theory and Related Topics*, ed. S. Kuroda, RIMS Publ. 242.

[16] J. Combes, P. Duclos and R. Seiler, Krein's formula and one dimensional multiple-well, J. Func. Anal., to appear.

[17] ______, Convergent expansions for tunneling, Comm. Math. Phys., to appear.

[18] N. Corngold, E. Harrell and B. Simon, The mathematical theory of resonances which have exponentially small widths, II, J. Math. Anal. Appl., to appear.

[19] E. B. Davies, Metastable states of symmetric Markov semigroups, I, Ann. Inst. H. Poincaré **35A** (1981), 149–171.

[20] ______, Metastable states of symmetric Markov semigroups, II, J. London Math. Soc., to appear.

[21] ______, JWKB and related bounds on Schrödinger eigenfunctions, Bull. London Math. Soc. **14** (1982), 273–284.

[22] ______, Double well Hamiltonians, Kings College preprint.

[23] ______, The twisting trick for double well Hamiltonians, Comm. Math. Phys. **85** (1982), 471–479.

[24] N. Fröman, The energy levels of double-well potentials, Ark. för Fysik **32** (1960), 79–97.

[25] E. Gildener and A. Patrascioiu, Pseudoparticle contributions to the energy spectrum of a one dimensional system, Phys. Rev. **D16** (1977), 425–443.

[26] J. Glimm and A. Jaffe, *Quantum Physics*, Springer-Verlag, Berlin & New York, 1981.

[27] L. Gross, Logarithmic Sobolev inequalities, Amer. J. Math. **97** (1976), 1061–1083.

[28] E. Harrell, On the rate of asymptotic eigenvalue degeneracy, Comm. Math. Phys. **60**, (1978), 73–95.

[29] ______, The band structure of a one dimensional periodic system in a scaling limit, Ann. Phys. **119** (1979), 351–369.

[30] ______, Double wells, Comm. Math. Phys. **75** (1980), 239–261.

[31] E. Harrell and B. Simon, The mathematical theory of resonances which have exponentially small widths, Duke Math. J. **47** (1980), 845–902.

[32] I. HERBST, Dilation analyticity in constant electric field, I: The two body problem, Comm. Math. Phys. **64** (1979), 279–298.
[33] G. JONA-LASINIO, F. MARTINELLI and E. SCOPPOLA, New approach to the semiclassical limit of quantum mechanics, I. Multiple tunnelings in one dimension, Comm. Math. Phys. **80** (1981), 223.
[34] M. KAC and C. THOMPSON, Phase transitions and eigenvalue degeneracy of a one dimensional anharmonic oscillator, Stud. Appl. Math. **48** (1969), 257–264.
[35] T. KATO, On the upper and lower bounds of eigenvalues, J. Phys. Soc. Japan **4** (1949), 334–339.
[36] L. LITHNER, A theorem of Phragmon-Lindelof type for second order elliptic operators, Ark. Math. **5** (1965), 281–285.
[37] S. MOLCHANOV, Diffusion processes and Riemannian geometry, Russian Math. Surveys **30** (1975), 1–63.
[38] M. PINCUS, Gaussian processes and Hammerstein integral equations, Trans. AMS **134** (1968), 193–216.
[39] A. M. POLYAKOV, Quark confinement and topology of gauge theories, Nucl. Phys. **B120** (1977), 429–458.
[40] M. REED and B. SIMON, *Methods of Modern Mathematical Physics, Vol.* IV. *Analysis of Operators*, Academic Press, 1977.
[41] W. REINHARDT, Complex coordinates in the theory of atomic and molecular structure and dynamics, Ann. Rev. Phys. Chem. **33** (1982), 223–255.
[42] J. ROSEN, Sobolev inequalities for weighted spaces and supercontractive estimates, Trans. AMS **222** (1976), 367–376.
[43] M. SCHILDER, Some asymptotic formulas for Wiener integrals, Trans. AMS **125** (1965), 63–85.
[44] B. SIMON, The theory of resonances for dilation analytic potentials and the foundations of time dependent perturbation theory, Ann. of Math. **97** (1973), 247–274.
[45] ______, Pointwise bounds on eigenfunctions and wave packets in $N$-body quantum systems, III. Trans. AMS **208** (1975), 317–329.
[46] ______, Semiclassical analysis of low lying eigenvalues, I. Non-degenerate minima: Asymptotic expansions, Ann. Inst. H. Poincaré **38** (1983), 295–307.
[47] ______, Instantons, double wells and large deviations, Bull. AMS **8** (1983), 323–326.
[48] ______, Large orders and summability of eigenvalue perturbation theory: A mathematical overview, Int. J. Quant. Chem. **21** (1982), 3–25.
[49] ______, Schrödinger semigroups, Bull. AMS **7** (1982), 447–526.
[50] ______, *Functional Integration and Quantum Physics*, Academic Press, 1979.
[51] ______, *Trace Ideals and Their Applications*, Cambridge University Press, 1979.
[52] T. SPENCER, The Lipatov argument, Comm. Math. Phys. **74** (1980), 273.
[53] E. C. TITCHMARSH, *Eigenfunction Expansions Associated with Second Order Differential Equations*, *Part* II, Oxford University Press, 1958.
[54] S. R. S. VARADHAN, On the behavior of the fundamental solution of the heat equation with variable coefficients, Comm. Pure Appl. Math. **20** (1967), 431–455.
[55] ______, Diffusion processes on a small time interval, Comm. Pure Appl. Math. **20** (1967), 659–685.
[56] ______, *Large Deviations and Applications*, CBMS Conference Proc., to be published by SIAM.
[57] E. WITTEN, Supersymmetry and Morse theory, J. Diff. Geo. **17** (1982), 661.

(Received September 27, 1983)