A general framework for discrete variable representation basis sets

Robert G. Littlejohn and Matthew Cargo
Department of Physics, University of California, Berkeley, California 94720

Tucker Carrington, Jr.
Département de Chimie, Université de Montréal, Case Postale 6128, Montréal, Québec H3C 3J7, Canada

Kevin A. Mitchell
Department of Physics, College of William and Mary, Williamsburg, Virginia 23185

Bill Poirier
Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409-1061

(Received 17 December 2001; accepted 7 March 2002)

A framework for discrete variable representation (DVR) basis sets is developed that is suitable for multidimensional generalizations. Those generalizations will be presented in future publications. The new axiomatization of the DVR construction places projection operators in a central role and integrates semiclassical and phase space concepts into the basic framework. Rates of convergence of basis set expansions are emphasized, and it is shown that the DVR method gives exponential convergence, assuming conditions of analyticity and boundary conditions are met. A discussion of nonorthogonal generalizations of DVR functions is presented, in which it is shown that projected δ-functions and interpolating functions form a biorthogonal basis. It is also shown that one of the generalized DVR proposals due to Szalay [J. Chem. Phys. **105**, 6940 (1996)] gives exponential convergence. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1473811]

I. INTRODUCTION

This paper is the first in a series that presents the results of our recent investigations into multidimensional discrete variable representation (DVR) basis sets. The basic ideas of the DVR method go back to the 1960's,[1,2] and the method has become widely popular especially since the pioneering work of Light and others in the 1980's.[3–11] A recent review has been given by Light and Carrington,[12] and another article we have found useful for general information is Baye and Heenen.[13]

There has long been felt a need for multidimensional DVR functions.[14] Cartesian products of one-dimensional DVR functions are easy to construct, but typically do not obey the boundary conditions needed on spaces, such as those occuring in molecular physics, which are not themselves Cartesian products of one-dimensional spaces. For these reasons we have been interested to find ways of constructing nontrivial (that is, non-Cartesian-product), multidimensional DVR basis sets. Our first step in doing this was to try to understand the known, one-dimensional DVR basis sets from as deep a standpoint as possible, in order to generalize them to higher dimensions. This paper presents the results of the first stage of this research, in which we axiomatize the DVR construction in a way which is not biased toward one-dimensional examples and identify some of the underlying principles.

In the process, we have developed several ideas. First, we have integrated the basic theory of DVR functions with phase space or semiclassical concepts, which we feel are an important part of understanding their properties and the principles of their construction. The phase space perspective on DVR functions has been explored previously by Poirier and Light.[15–17] In our axiomatization of the DVR construction, we have promoted projection operators to a central role, and emphasized the phase space meaning of these. We have also paid considerable attention to the rates of convergence of different basis set methods, and emphasized the favorable case of exponential convergence. In particular, we have demonstrated that the DVR method provides exponential convergence under certain circumstances (analyticity of the potential, proper attention to boundary conditions, etc.). Although it is well known that the DVR method often gives rapid convergence, we believe our argument for the exponential convergence of the DVR method is new. Another novel element is our new proof of the Darboux–Christoffel formula, a basic result in the theory of one-dimensional, orthogonal polynomial DVR functions. Finally, we present a theory of DVR functions which are generalized by relaxing the usual, orthogonality conditions. The idea of such generalized DVR functions was first put forward by Light *et al.*[18] and was later developed by Szalay,[19] but we have presented several new ideas including the duality between the set of projected δ-functions and the set of interpolating functions and a proof of exponential convergence with a certain form for the matrix elements of the potential energy.

II. PROJECTION OPERATORS, PHASE SPACE, AND EXPONENTIAL CONVERGENCE

In our axiomatization of the DVR construction we have promoted projection operators to a central role, a step that turns out to be important when constructing multidimensional examples. In this we were strongly influenced by the treatment of Baye and Heenan.[13] Moreover, projection op-

erators have a geometrical interpretation in phase space which is very suggestive, not only for understanding the efficiency and rate of convergence of known DVR basis sets, but also for designing new ones. In favorable circumstances DVR basis sets lead to exponential convergence in the calculation of eigenvalues and eigenfunctions. In this section we discuss the conditions under which exponential convergence holds and the factors which can destroy it.

We begin with projection operators. Let $\mathcal{H}$ be a Hilbert space of wave functions, and let $\{|n\rangle, n=0,\ldots,N-1\}$ be a truncated, orthonormal basis. Let

$$P=\sum_{n=0}^{N-1}|n\rangle\langle n| \tag{2.1}$$

be the projection operator onto the subspace $\mathcal{S}$ spanned by the truncated basis set, so that $\mathcal{S}=P\mathcal{H}$. In practice P is specified by the truncated basis $\{|n\rangle\}$. The stationary phase logic of this paper requires that at least part of the basis be a truncated spectral basis, but it may contain additional states that do not fit this description.[20] We prefer to emphasize P itself as the primary object because the basis which spans $\mathcal{S}$ is not unique and because in multidimensional problems it is often much less obvious what is to be regarded as a privileged choice of the basis $\{|n\rangle\}$. Moreover, P may be interpreted in terms of a region of phase space, which is independent of the choice of basis spanning $\mathcal{S}$. The phase space interpretation of P plays an important role in this paper, and also in the problem of basis set optimization, that is, the problem of choosing a P which is efficient for a particular class of wave functions one wishes to find. This problem has been considered by several authors.[4,16,17,21–28] The phase space interpretation of P is a natural outcome of the Wigner–Weyl formalism,[29–31] which has recently been specifically elaborated upon by Poirier.[15] Here we shall merely use a simple model problem to present the intuitive idea.

Suppose we are trying to solve the Morse oscillator in a harmonic oscillator basis. Figure 1(a) illustrates the Morse potential (above) and the phase space of the Morse oscillator (below). In the phase space diagram, a set of nested classical orbits is drawn, having actions $I_n=(1/2\pi)\oint p\,dx=n\hbar$, $n=1,2,\ldots$ (the circle in the figure will be explained momentarily). Thus, the area of the nth orbit is $nh=n(2\pi\hbar)$ (the orbit contains n Planck cells of phase space area). These are not the quantized orbits of WKB theory, which satisfy $I_n=(n+1/2)\hbar$, $n=0,1,\ldots$; the quantized orbits are half way between the orbits drawn in the figure. We have drawn the figure this way because we wish to imagine the quantized orbits at the centers of annular strips, each of which contains one Planck cell ($h=2\pi\hbar$) of phase space area.

Figure 1(b) illustrates a harmonic oscillator potential (above) and phase space (below). The potential is the harmonic approximation to the Morse potential at the bottom of the well. The phase space picture of the harmonic oscillator also contains a set of nested orbits, in this case circles, having actions $I_n=(1/2\pi)\oint p\,dx=n\hbar$. The number of harmonic oscillator orbits and Morse oscillator orbits drawn (10) is the same.

The circle in the Morse oscillator phase space is the last (tenth) harmonic oscillator orbit. This circle completely encloses the first three of the Morse oscillator orbits, but the fourth extends beyond it a small amount near the x-axis. This means that if we solve the Morse oscillator in a harmonic oscillator basis truncated at the tenth basis state ($n=9$ in the usual numbering $n=0,1,\ldots$), then the first three Morse oscillator states ($n=0,1,2$) will be reasonably well converged, but the fourth ($n=3$) will have some qualitative error in the eigenfunction $\psi_3(x)$ near the right turning point, and the qualitative disagreement will get worse for higher states $n=4,5,\ldots$.

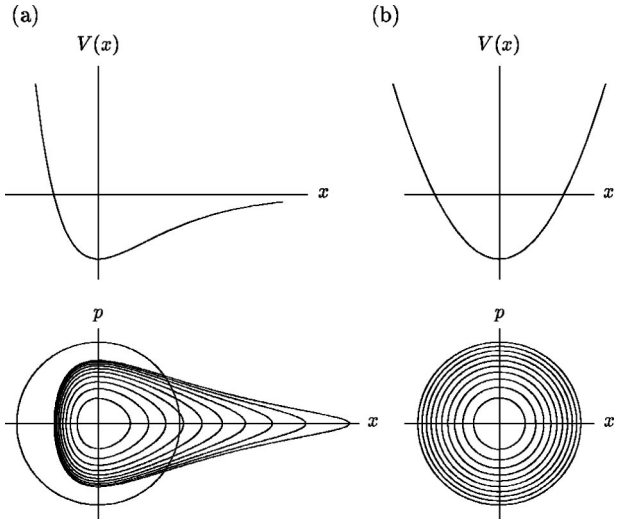

FIG. 1. In (a), the potential (above) and orbits in phase space (below) for the Morse oscillator. The orbits have actions $I_n=n\hbar$. The circle in the phase space diagram is the outer harmonic oscillator orbit in (b). In (b), potential and phase space orbits for a harmonic oscillator. The potential is the harmonic approximation to the Morse potential.

To be more precise, consider the error in the mth Morse oscillator eigenstate $\psi_m(x)=\langle x|m\rangle$ when computed in a truncated basis of N harmonic oscillator eigenstates $\{|n\rangle, n=0,\ldots,N-1\}$ as N increases. This error can be estimated by the matrix element $\langle N|m\rangle$, the coefficient of the first neglected term in the expansion of the exact Morse oscillator eigenstate in the harmonic oscillator basis. This matrix element in turn can be approximated by using semiclassical (WKB) wave functions and the stationary phase approximation. The matrix element $\langle N|m\rangle$ then becomes a sum of integrals of the form,

$$\int dx\sqrt{\cdots}\exp\left\{\frac{i}{\hbar}[\pm S_N^{\mathrm{HO}}(x)\pm S_m^{\mathrm{MO}}(x)]\right\}, \tag{2.2}$$

where $S=\int p\,dx$, where the ellipsis indicates the usual WKB amplitude factors and where the superscripts HO and MO refer to the harmonic and Morse oscillators, respectively. The integrals are summed over the choices of sign to get the matrix element $\langle N|m\rangle$. The stationary phase condition can be satisfied only when the two signs are opposite, so that the stationary phase points are the roots of $(d/dx)[S_N^{\mathrm{HO}}-S_m^{\mathrm{MO}}(x)]=0$, or $p_N^{\mathrm{HO}}(x)=p_m^{\mathrm{MO}}(x)$, where $p=p(x)$ is the momentum function defined by the classical orbits. Thus, the stationary phase points are represented geometrically by the

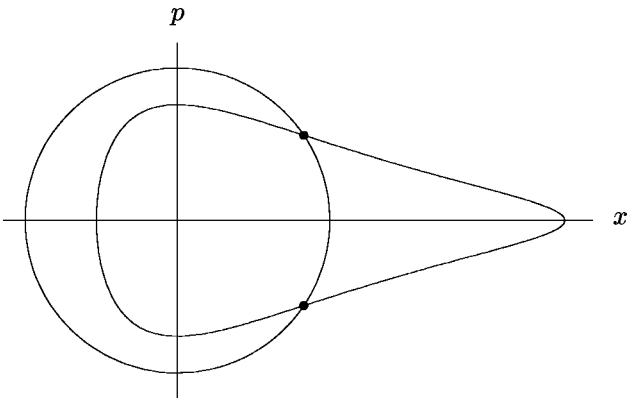

FIG. 2. The integral for the computation of the matrix element $\langle N|m\rangle$, the Nth expansion coefficient in the expansion of the mth Morse oscillator state in a harmonic oscillator basis, is dominated by contributions from the stationary phase points, which are geometrically the intersections of the Nth harmonic oscillator orbit with the mth Morse oscillator orbit.

intersection of the Nth harmonic oscillator quantized orbit with the mth quantized Morse oscillator orbit, as illustrated in Fig. 2. (See Littlejohn[32,33] for a discussion of the multidimensional case.)

As N increases (for fixed m) there comes a value of N where the Nth circle covers the mth Morse orbit, and as N increases beyond that value there will be no more intersections of the two orbits. This means that there are no real stationary phase points in the computation of the matrix element $\langle N|m\rangle$; the stationary phase points have effectively moved off the real axis and become complex. The complex stationary phase points correspond to actions S in the exponent $e^{iS/\hbar}$ which have a positive imaginary part; the matrix element itself is exponentially decreasing in the imaginary part of S. Thus, if we describe the error in terms of its dependence on N, we expect this error to be $O(1)$ (that is, large) when the two orbits have real intersections, and to decrease exponentially (as e^{-cN} for some constant c) as the area covered by the basis states expands beyond the area occupied by the Morse eigenstates. [To be more precise, the error will go as $e^{-c(N-N_0)}$ when N is above but near N_0, the number of basis states which just cover the mth Morse oscillator state.]

The harmonic oscillator basis is an example of an *analytic, spectral basis*, that is, a basis consisting of the analytic eigenfunctions of some operator. Of course, the Morse oscillator eigenfunctions form another analytic, spectral basis. Exponential convergence occurs whenever one analytic, spectral basis is expanded in terms of another, although it only sets in after the phase space area of the unknown eigenfunctions (or volume in higher dimensions) has been covered by the basis, and if the constant c is small the convergence may not be satisfactory. That is, both the basis set size N_0 above which exponential convergence sets in and the constant c determining the rate of convergence depend on the basis, and determine its efficiency. Nevertheless, in favorable cases the exponential convergence is dramatic, and in general it seems to be a highly desirable property for a basis set to have. Bases which are not analytic generally do not give exponential convergence; these include splines (or any kind of piecewise fits) and ordinary (multiresolution) wavelets,[34–36] which typically have compact support and so are not analytic.

Exponential convergence is also destroyed if the basis functions and the functions to be expanded do not satisfy the same boundary conditions. In current practice, this condition is not always met, although boundary conditions can be ignored if the physical wave function has negligible amplitude at the boundaries. In a recent paper[37] we have examined the boundary conditions satisfied by physical wave functions in the internal space of the 3-body problem. We also point out that particularly careful attention has been given to boundary conditions at collinear shapes in the treatment of the H_3^+ molecule in Ref. 38.

Exponential convergence will not hold either if the wave functions to be expanded are themselves nonanalytic, even if the basis is analytic. This occurs in the case of electronic wave functions, due to the cusp singularities.[39] Such singularities in configuration space give rise to long-range tails in momentum space. Wave functions which are analytic in configuration space normally correspond to momentum space wave functions which fall off exponentially as $p\to\infty$; this can typically be interpreted as a tunnelling into the classically forbidden region in momentum space (beyond the momentum turning points). Nuclear wave functions in the Born–Oppenheimer approximation also have cusps at the coincidence of two nuclei, but of course the wave function is strongly suppressed by the potential there. Another issue is that in practice the potential energy surface is usually nonanalytic, being made up of piecewise analytic fits to electronic structure data points. These nonanalyticities (where the pieces fit together) will destroy exponential convergence at some level of accuracy regardless of the basis, but of course no physics can depend on these nonanalyticities. As for other types of singularities (at conical intersections or singularities due to body frame,[40–42] etc.) it seems to us that it should be possible to handle these without loss of exponential convergence assuming other problems are solved (a tall order). To summarize a somewhat complex situation, it seems to us that as a first pass one should concentrate on analytic bases for nuclear wave functions in molecular problems, which is the philosophy we have adopted in our studies of DVR bases.

The above analysis of convergence is based on WKB theory, which applies to analytic eigenfunctions and which hopefully has intuitive advantages for chemists and physicists. It is necessary background for our discussion of the rate of convergence of the DVR method below. See Gottlieb and Orszag[43] for a more standard treatment of convergence rates in expansions in orthogonal functions, in which smooth (rather than analytic) functions are emphasized, and the effects of boundary conditions are discussed. As far as we know, the rates of convergence of the DVR method (by which we mean the errors incurred when the standard diagonal approximation is made for the potential energy and eigenvalues of the truncated matrix are computed) have not been considered in the mathematical literature.

To return to the projection operator P, in the case of analytic bases this operator is associated with a region of

phase space. Of course, the range of unknown eigenstates one wishes to find is also associated with a region of phase space, and another projection operator. In general, the Weyl transform of a projection operator is a function which to some approximation is the characteristic function of a region of phase space (zero outside, unity inside). This is only an approximation, the nature of which has been recently explored by Poirier,[15] including some numerical examples. See also Berry,[44] who studies the Weyl transforms of projection operators for single states, and shows that they are smoothed δ-functions concentrated on the quantizing orbit, plus oscillations. The smoothing is enough to spread the δ-function over the width of the annular strips illustrated in Fig. 1, producing an approximation to the characteristic function of the annular strip. We note that in this paper we never use these approximations for any quantitative purpose; our quantitative conclusions rely on the stationary phase argument given above for exponential convergence.

III. A GENERAL FRAMEWORK FOR DVR

We shall now present a framework for defining and discussing DVR functions which is not biased toward one-dimensional examples or the special case of orthogonal polynomials.

Let M be the manifold (or space) upon which our wave functions live. In model problems (and some real problems) M may be the Euclidean space $\mathbb{R}^n$, but in molecular quantum mechanics M is typically one of the internal spaces of the n-body problem, with nontrivial topology. We have examined the topology and structure of these spaces in a series of earlier papers.[41,42,45] The only mathematical manifolds which are Cartesian products of one-dimensional manifolds are planes, cylinders and tori (generally, products of circles and lines). Unfortunately, these do not include the internal spaces primarily of interest in molecular quantum mechanics. This is one reason for being interested in non-Cartesian product, multidimensional DVR bases.

Let us denote the Hilbert space of square-integrable wave functions on M by $\mathcal{H}=L^2(M)$ (we assume M has a metric). Let P be a projection operator on $\mathcal{H}$, and let $\mathcal{S}=P\mathcal{H}$ be the subspace upon which P projects. Let $\{x_\alpha, \alpha=0,\ldots,N-1\}$ be a set of N grid points on M. If M is d-dimensional, then each x_α is represented by a set of d coordinates. Finally, we introduce the projected δ-functions concentrated at the grid points, $\Delta_\alpha(x)=P[\,\delta(x-x_\alpha)]$, which we denote with a capital Δ to emphasize that they came from δ-functions. This is more convenient in Dirac notation, $|\Delta_\alpha\rangle=P|x_\alpha\rangle$, so that $\Delta_\alpha(x)=\langle x|\Delta_\alpha\rangle$.

With these definitions we have the following theorem:

$$\langle\Delta_\alpha|\Delta_\beta\rangle=\Delta_\beta(x_\alpha)=\Delta_\alpha(x_\beta)^*, \tag{3.1}$$

which says that the matrix of scalar products of the projected δ-functions (the overlap matrix) is given by evaluating the projected δ-functions at each others' grid points. The proof is trivial; we have

$$\langle\Delta_\alpha|\Delta_\beta\rangle=\langle x_\alpha|P^\dagger P|x_\beta\rangle=\langle x_\alpha|P|x_\beta\rangle=\langle x_\alpha|\Delta_\beta\rangle=\langle\Delta_\alpha|x_\beta\rangle, \tag{3.2}$$

where we use $P^\dagger P=P^2=P$ and $P|x_\alpha\rangle=|\Delta_\alpha\rangle$.

This theorem has an immediate corollary, which says that the set of projected δ-functions, $\{\Delta_\alpha(x),\alpha=0,\ldots N-1\}$ is orthogonal if and only if

$$\Delta_\alpha(x_\beta)=K_\alpha\delta_{\alpha\beta}, \tag{3.3}$$

where $K_\alpha=\langle\Delta_\alpha|\Delta_\alpha\rangle$ is the squareed norm of the projected δ-function (in general they are not normalized). If in addition the projected δ-functions do not vanish identically, then $K_\alpha>0$ and we can define a normalized (hence orthonormal) version of these functions,

$$|F_\alpha\rangle=\frac{1}{\sqrt{K_\alpha}}|\Delta_\alpha\rangle, \tag{3.4}$$

so that $\langle F_\alpha|F_\beta\rangle=\delta_{\alpha\beta}$.

We now define a *DVR set* as the combination of a projection operator P plus a set of N grid points $\{x_\alpha\}$ such that Eq. (3.3) holds with all $K_\alpha>0$. In view of the theorem (3.1), we are requiring that the projected δ-functions vanish at each others' grid points and that they do not vanish at their own grid points. All the standard examples also satisfy another condition, which is $N=\dim\mathcal{S}$, that is, the number of grid points (hence the number of projected δ-functions) is equal to the dimensionality of the space upon which P projects. (This space is sometimes infinite-dimensional, however.) Since these functions are orthogonal and nonvanishing, they are necessarily linearly independent, and hence span the subspace $\mathcal{S}$ (in the finite dimensional case). The $\{F_\alpha(x)\}$ are then an orthonormal basis on this subspace.

The following will help translate the notation of this paper into that used by other authors. Most authors define the subspace $\mathcal{S}$ as that spanned by an orthonormal set $\{\phi_n\}$, that is, the FBR. Thus, $P=\Sigma_n|\phi_n\rangle\langle\phi_n|$. P is denoted I_N by Light *et al.*[18] Most authors label grid points by α and call the grid points themselves x_α, as we do. The functions $F_\alpha(x)$ are called $\theta_\alpha(x)$ by Light and Carrington.[12] The squared norm of the projected δ-functions, K_α, turns out to be the inverse of the weights in the associated quadrature formula, denoted w_α or ω_α by most authors.

We make one remark here about the case $N\neq M\equiv\dim\mathcal{S}$. Since the functions $\Delta_\alpha(x)$, $\alpha=0,\ldots,N-1$, lie in the subspace $\mathcal{S}$, at most M of them are linearly independent. Thus the overlap matrix $\langle\Delta_\alpha|\Delta_\beta\rangle$, which is necessarily nonnegative definite, satisfies

$$\mathrm{rank}\langle\Delta_\alpha|\Delta_\beta\rangle\leqslant M. \tag{3.5}$$

Thus, if $N>M$ (the most interesting case after $M=N$, since it allows more grid points for quadrature purposes than basis functions), the matrix $\langle\Delta_\alpha|\Delta_\beta\rangle$ is a singular matrix. In fact, normally (apart from some deliberate or perverse choices of grid points) the rank is precisely M, and $\langle\Delta_\alpha|\Delta_\beta\rangle$ has precisely M positive and $N-M$ zero eigenvalues. Then Eq. (3.3) (with $K_\alpha>0$) is impossible. Henceforth in this article we shall deal only with the case $N=M$ (but in future articles we shall show how understanding the case $N>M$ is important for constructing DVR sets on multidimensional spaces).

DVR functions satisfy two properties simultaneously: The first is orthogonality, $\langle\Delta_\alpha|\Delta_\beta\rangle=K_\alpha\delta_{\alpha\beta}$ or $\langle F_\alpha|F_\beta\rangle$

$=\delta_{\alpha\beta}$; and the second is the interpolation property, $\Delta_\alpha(x_\beta)=K_\alpha\delta_{\alpha\beta}$, that is, each $\Delta_\alpha(x)$ vanishes at all grid points except its own.

We will call a set of N linearly independent functions $\{L_\alpha(x)\}$ which span some given function space *interpolating functions* if they satisfy $L_\alpha(x_\beta)=\delta_{\alpha\beta}$ for a given set of N grid points $\{x_\alpha\}$. These functions are unique if they exist. In the case of a DVR set, a set of interpolating functions which span the space $\mathcal{S}$ is given by

$$L_\alpha(x)=\frac{1}{K_\alpha}\Delta_\alpha(x). \tag{3.6}$$

For a DVR set, all three functions, Δ_α, F_α, and L_α, are proportional.

The orthogonality and interpolating properties of DVR functions give rise to two distinct ways to expand a function which belongs to $\mathcal{S}$. That is, if $\psi\in\mathcal{S}$, so that an expansion of the form

$$\psi(x)=\sum_\alpha c_\alpha F_\alpha(x) \tag{3.7}$$

exists, then the expansion coefficients can be determined either by orthogonality or by setting $x=x_\beta$ in Eq. (3.7) and using Eq. (3.3). That is,

$$c_\alpha=\int dx\, F_\alpha^*(x)\psi(x)=\frac{1}{\sqrt{K_\alpha}}\psi(x_\alpha). \tag{3.8}$$

Thus, when $\psi\in\mathcal{S}$, we can write

$$\psi(x)=\sum_\alpha \frac{1}{\sqrt{K_\alpha}}\psi(x_\alpha)F_\alpha(x). \tag{3.9}$$

If ψ does not belong to $\mathcal{S}$, then the two (nominal) ways of determining the expansion coefficients do not give the same results, and the errors in the two expansions are different. As we have seen in Sec. II, if ψ lies well within the region of phase space covered by P, then the errors in the orthogonality expansion ($c_\alpha=\langle F_\alpha|\psi\rangle$) are exponentially small. We now show that the same is true for the interpolation expansion ($c_\alpha=(1/\sqrt{K_\alpha})\psi(x_\alpha)$), under certain assumptions that are common in practice.

First we break ψ into a part ψ_1 lying in $\mathcal{S}$, and another ψ_2 orthogonal to $\mathcal{S}$, $\psi=\psi_1+\psi_2$. Then the error in the interpolating expansion is

$$\begin{aligned}\mathrm{err}(x)&=\psi(x)-\sum_\alpha \frac{1}{\sqrt{K_\alpha}}\psi(x_\alpha)F_\alpha(x)\\ &=\psi_2(x)-\sum_\alpha \frac{1}{\sqrt{K_\alpha}}\psi_2(x_\alpha)F_\alpha(x),\end{aligned} \tag{3.10}$$

since ψ_1 has an exact interpolating expansion. Let $\mathrm{err}(x)=\langle x|\epsilon\rangle$. Then the squared norm of the error is given by

$$\langle\epsilon|\epsilon\rangle=\langle\psi_2|\psi_2\rangle+\sum_\alpha \frac{1}{K_\alpha}|\psi_2(x_\alpha)|^2, \tag{3.11}$$

since the two major terms on the right-hand side of Eq. (3.10) are orthogonal. The first term of Eq. (3.11) was shown to be exponentially small by the argument in Sec. II. The second term is the squared norm of ψ_2, computed by the quadrature approximation, which need not be a good one since ψ_2 lies outside $\mathcal{S}$. The magnitude of the second term depends on the values of the function ψ_2 at the grid points. Although ψ_2 has a small norm, this does not in principle prevent ψ_2 from taking on large values at certain points. In many practical circumstances, however, the values of ψ_2 will be exponentially small.

Suppose, for example, that $\mathcal{S}$ is spanned by an analytic, spectral basis $\phi_n(x)$, truncated at $n<N$. Then $\psi_2(x)=\sum_{n=N}^\infty a_n\phi_n(x)$, where, according to the arguments in Sec. II, a_N is of order e^{-cN} and the terms a_n decrease exponentially after that. Now note that

$$|\psi_2(x)|\leqslant\sum_{n=N}^\infty |a_n||\phi_n(x)|. \tag{3.12}$$

This sum will be dominated by the first few terms unless the maximum value of $\phi_n(x)$ diverges fast enough to counteract the exponential decay of the a_n. In fact, this maximum value may diverge; for example, the maximum value of the harmonic oscillator eigenfunctions goes as $n^{1/12}$, and the maximum value of the Y_{lm}'s on the sphere diverges as $l^{1/6}$, but these are very mild divergences that are completely overwhelmed by the exponential decay of the coefficients a_n. It is also possible that the eigenfunctions $\phi_n(x)$ themselves diverge at some points x, for example, the radial eigenfunctions of the Dirac hydrogen atom do this, but these divergences are due to the singularity of the potential at $r=0$, and here we are assuming that all potentials are analytic.

A more general argument is the following. Suppose $\psi_2(x)$ reaches a maximum value of H in a lobe of width Δx. Then the contribution of this lobe to the squared norm $\nu=\langle\psi_2|\psi_2\rangle$ is of order $H^2\Delta x\leqslant\nu$. We are assuming that $\psi(x)$ is an eigenfunction in an analytic potential, so its Fourier transform dies off exponentially beyond the classically allowed region in momentum space. Similarly, the Fourier transform of ψ_2, which is the truncation of ψ outside the region of phase space covered by $\mathcal{S}$, dies off exponentially outside some momentum bound P_0 of this region. This means that the maximum momenta in the Fourier transform of the one lobe containing the maximum of $\psi_2(x)$ is of the order of P_0, or $\Delta x>\hbar/P_0$. Thus, $\hbar H^2/P_0<\nu$, or $H<(P_0\nu/\hbar)^{1/2}$. The cutoff momentum P_0 is a slowly increasing function of N, the size of the truncated basis, but this is completely dominated by the exponential decay of the squared norm ν. Thus, the maximum value of $\psi_2(x)$ decays exponentially with N, and the squared norm of the error of the interpolating expansion (3.11) is exponentially small.

IV. STANDARD DVR SETS

In this section we show how certain standard DVR sets, namely, sinc functions and DVR sets based on orthogonal polynomials, fit into the general formalism we have presented, and we discuss convergence issues and phase space interpretations of these functions.

Sinc functions[46,47] are in many ways the simplest example of a DVR basis and also an instructive one, one which is worth examination because of the many fundamental

points it illustrates. Sinc functions do not satisfy the boundary conditions required in many practical problems, but when the boundary conditions are right (on the spaces $\mathbb{R}$ or $\mathbb{R}^n$), sinc functions produce exponential convergence according to the same principles of phase space area which apply to any other DVR basis. Moreover, the long-range tails of sinc functions are not a problem, as we explain.

We begin by showing how sinc functions fit into the DVR formalism presented in Sec. III. For this we choose $M=\mathbb{R}$ (we work with one-dimensional wave functions), we choose the projection operator to be

$$P=\int_{-p_0}^{+p_0} dp\,|p\rangle\langle p|, \tag{4.1}$$

where $|p\rangle$ is a momentum eigenstate (a plane wave) and p_0 is a momentum cutoff, and we choose the grid points to be $x_n=na$, where $a=\pi\hbar/p_0$ is the lattice spacing. For sinc functions we use n instead of α to label the grid points; n is an arbitrary integer. The space $\mathcal{S}$ upon which P projects is infinite-dimensional, and consists of band-limited functions, that is, those whose momentum-space wave function vanishes outside $|p|\leqslant p_0$. The DVR property follows from

$$\Delta_n(x)=\langle x|P|x_n\rangle=\int_{-p_0}^{p_0}\frac{dp}{2\pi\hbar}e^{ip(x-x_n)/\hbar}$$
$$=\frac{1}{\pi}\frac{\sin[\pi(x-na)/a]}{x-na}, \tag{4.2}$$

which shows that $\Delta_n(x_m)=\delta_{nm}/a$, so that $K_n=a$ (independent of n).

To prove that the functions $\Delta_n(x)$ are complete on $\mathcal{S}$ we cannot simply count linearly independent functions, since $\mathcal{S}$ is infinite-dimensional. Instead we note that the set of functions in momentum space,

$$f_n(p)=\frac{1}{\sqrt{2p_0}}e^{-in\pi p/p_0} \tag{4.3}$$

is obviously orthonormal and complete for functions defined on the interval $-p_0\leqslant p\leqslant p_0$, since it is just a Fourier series basis on that interval. Transforming back to x-space, we find that the Fourier transform of $f_n(p)$ is $\sqrt{a}\Delta_n(x)=F_n(x)$, so the set $\{F_n(x)\}$ is also orthonormal and complete on $\mathcal{S}$.

A Fourier series basis on an interval converges slowly for functions which are not periodic (in fact, the function should be analytic and periodic for exponential convergence). Thus, we should at least require $\phi(p_0)=\phi(-p_0)$ for the momentum space wave functions whose x-space counterparts we wish to expand in a sinc function basis. In applications of sinc DVR functions in quantum mechanics, one will probably be dealing with momentum space wave functions $\phi(p)$ which are very small at $p=\pm p_0$, so that $\phi(p_0)\approx\phi(-p_0)\approx 0$. The general idea is illustrated in Fig. 3, in which we are thinking of solving a one-dimensional oscillator in a sinc function basis. The highest energy eigenfunction desired corresponds to a classical orbit which is sketched as an oval in the diagram, and the region of phase space covered by the sinc functions is the band $-p_0\leqslant p\leqslant p_0$ centered on the x-axis. This band has infinite area, corresponding to

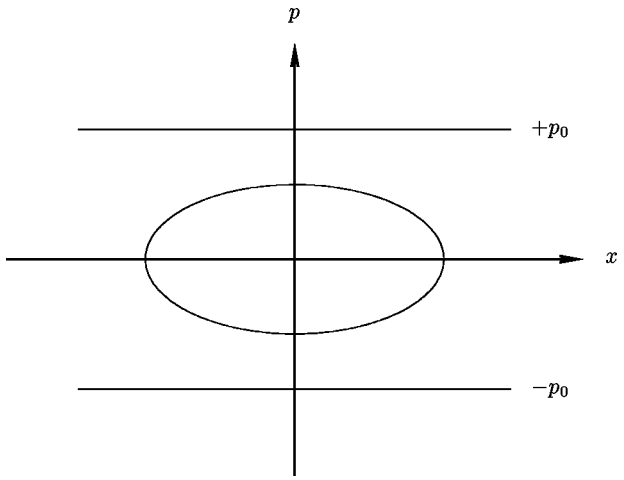

FIG. 3. Phase space diagram for finding the eigenfunctions and eigenvalues of a one-dimensional oscillator in a basis of sinc functions. The desired, unknown eigenfunctions occupy the oval region of phase space, and the sinc functions occupy the band $-p_0\leqslant p\leqslant p_0$ centered on the x-axis.

the infinite sinc function basis. We have chosen p_0 to lie well outside the classically allowed momentum values for the maximum energy desired, so that, if the potential is analytic, the momentum space wave function will be very small and decreasing exponentially at $p=\pm p_0$. The parameter p_0, which is specified by the spacing of the lattice of sinc functions through $p_0=\pi\hbar/a$, is a convergence parameter, and we shall have exponential convergence in this parameter for analytic eigenfunctions.

There is another convergence parameter in the use of sinc DVR functions, which is the truncation size of the basis (necessary since the basis is infinite). That is, we must decide how many lattice points in x-space to include in the basis. It is logical that we should go out beyond the turning points in x-space, but in view of the long-range tails of the sinc functions, one might worry that it would be necessary to go a very long way. In fact, this is not so, since the expansion coefficients of a band-limited wave function ψ in a sinc function basis are just proportional to the value of the wave function at the grid points, as shown by Eq. (3.9). Since ψ is decreasing exponentially in the classically forbidden region (now in x-space), the convergence is exponential in this truncation parameter, too. To be more precise about this, we must worry about the fact that the unknown eigenfunctions are not truly band limited, so the expansion (3.9) is not strictly valid. However, the amount by which ψ differs from a band limited function is related to the exponentially small value of the momentum space wave function at $p=\pm p_0$. Thus, if we fix p_0 and consider the error as we enlarge the number N of basis functions (that is, lattice points) in the truncated basis, we find a convergence which looks exponential up to the point we reach the errors due to the truncation of the momentum space wave function. Beyond that point, adding further basis functions does not help since even the infinite basis cannot represent momentum values $|p|>p_0$.

If, however, we increase p_0 at the same time we increase the truncation size of the basis, we obtain convergence which is overall exponential. In practice this works quite easily, and

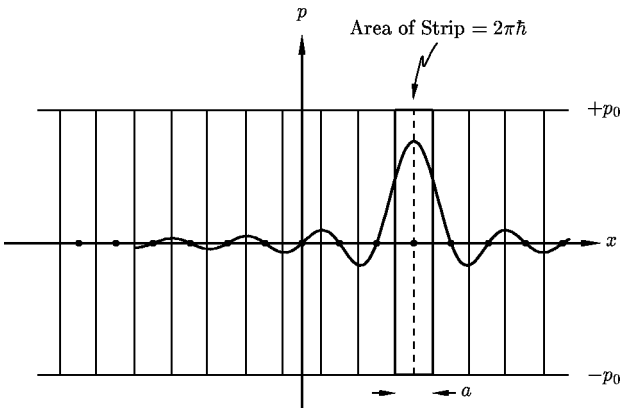

FIG. 4. Phase space diagram for the sinc DVR basis, illustrating the Planck cells (vertical strips) associated with each basis function.

exactly as predicted by the theory. The final result is some rectangle in phase space, bounded in the x-direction by the truncation of the basis and in the p-direction by the momentum cutoff p_0. This rectangle completely encloses the region containing the unknown eigenstates, and the ratio of basis states to the number of eigenstates found is the ratio of the area of the rectangle to the area of the classical orbit bounding the eigenstate region.

There remain some worries about the long-range tails of the sinc functions. The argument given above explains why the expansion coefficients of a wave function ψ in a sinc function basis associated with grid points well into the classically forbidden region must be exponentially small. This was based on the interpolation property for computing expansion coefficients [the final expression in Eq. (3.8)]. But we should also be able to use orthonormality to compute these expansion coefficients [the center expression in Eq. (3.8)]. How can this integral be exponentially small, when the long range tail of the sinc function, centered out in the classically forbidden region, overlaps substantially with the wave function ψ in the classically allowed region? To understand this, we express the overlap integral in terms of the complex integrals,

$$\int dx \frac{e^{\pm i\pi x/a}}{x-na}\psi(x), \tag{4.4}$$

where the grid point $x_n=na$ is in the classically forbidden region for $\psi(x)$. Under these circumstances, the fraction $1/(x-na)$ can be regarded as slowly varying on the scale of the exponential throughout the classically allowed region, where ψ is large. Thus the function $f(x)=\psi(x)/(x-na)$ has approximately the same momentum content as $\psi(x)$ itself, and the integral can be estimated by the Fourier transform of ψ evaluated at $p=\pm p_0$. But by construction, $\phi(p)$ is exponentially small at these (momentum cutoff) values, so the apparently dominant contribution to the integral (coming from the classically allowed region for ψ) is, in fact, exponentially small.

We now use Fig. 4 to make some points regarding the phase space perspective on the sinc functions. This is a hybrid figure, with a sinc function centered on a certain grid point superimposed on a phase space diagram. The band $-p_0\leqslant p\leqslant p_0$ in phase space covered by the sinc functions is shown, and the grid points $x=na$ are shown as dots on the x-axis. In a diagram like this, it is natural to associate each basis state with a vertical strip of width a, centered on the grid points. Since the height of each strip is $2p_0=2\pi\hbar/a$, the area per strip is $2\pi\hbar$, a single Planck cell. In this way, the infinite area of the band is divided into Planck cells, one for each basis state. This association is not merely a matter of making the area come out right, for the Lagrangian manifold[33] associated with the delta function $\delta(x-x_n)$ centered on grid point x_n is the vertical line in phase space given by $x=x_n$. When the projection operator P is applied to the δ-function to create $\Delta_n(x)$, the Lagrangian manifold is effectively truncated at $p=\pm p_0$, giving a finite line segment such as the vertical dotted line in the figure. The vertical strip is centered on this line much as the annular strips in Fig. 1 are centered on the quantizing orbits. This picture can be made more quantitative by computing the Wigner function of the sinc function, which turns out to be approximately the characteristic function of the vertical strip. The approximation is rather crude, however, because of the large amount of ringing due to the discontinuous cutoffs. Nevertheless, a picture like this captures some important semiclassical features of the sinc function basis.

Now we make some remarks about orthogonal polynomial DVR functions. Let $\{q_n(x),n=0,1,\ldots\}$ be a set of real, one-dimensional polynomials, where q_n is of degree n, that is orthonormal on an interval $[a,b]$ with respect to a weighting function $\rho(x)>0$, $\int_a^b\rho(x)q_n(x)q_m(x)=\delta_{nm}$. Let $\phi_n(x)=\sqrt{\rho(x)}q_n(x)$ be the weighted polynomial functions that are orthonormal in the usual sense, and let $P_N=\Sigma|\phi_n\rangle\langle\phi_n|$. Then, as is well known[2,13] a DVR set results if the grid points are chosen to be the roots of $q_N(x)$ (the first polynomial omitted from the projection operator). These facts follow from the Darboux-Christoffel formula, a standard result in the theory of one-dimensional, orthogonal polynomials.[48,49]

Unlike sinc DVR functions, the grid points of orthogonal polynomial DVR functions are spaced nonuniformly. This has a simple semiclassical interpretation. Consider the WKB approximation to $\phi_N(x)$,

$$\phi_N(x)=\left(\frac{2}{\pi}\frac{\partial^2 S}{\partial I\partial x}\right)^{1/2}\cos[S(x,I)/\hbar-\pi/4], \tag{4.5}$$

where $I=(N+1/2)\hbar$ is the action of the quantizing orbit for the state ϕ_N and where $S=\int p\,dx$ is measured from the left turning point. In the WKB approximation, the zeros of ϕ_N occur when the argument of the cosine is $(\alpha+1/2)\pi$, so the spacing Δx between the roots is given by $\Delta S/\hbar=(1/\hbar)\times(\partial S/\partial x)\Delta x=\pi$, or $\Delta x=\pi\hbar/p(x,I)$, where $p(x,I)=\partial S/\partial x$ is the local momentum of the first neglected state. This is identical to the spacing rule for sinc functions except that the constant spacing (a above) has been replaced by the variable Δx, and the constant momentum cutoff (p_0 above) has been replaced by the local momentum $p(x,I)$. Obviously the result is open to the same interpretation, that the DVR state occupies a vertical strip in phase space containing one Planck cell ($2\pi\hbar$) of area.

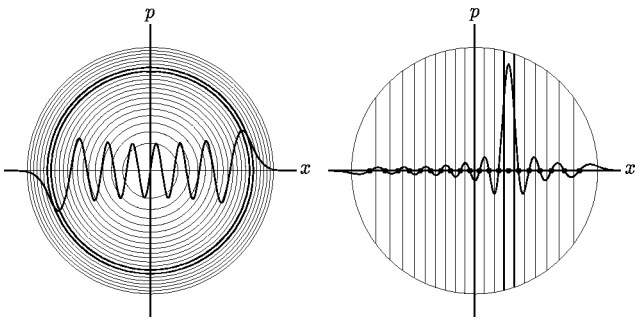

FIG. 5. Illustrations of the phase space area occupied by the first 20 states of the harmonic oscillator. On the left, the annular strips of equal area are occupied by the harmonic oscillator eigenstates $\phi_n(x)$, $n=0,\ldots,19$. The wave function plotted is $\phi_{13}(x)$, which corresponds to the annular strip bounded by heavy lines. On the right, the same area is divided into 20 vertical strips of equal area, occupied by the DVR states. The dots on the x-axis are the zeroes of $\phi_{20}(x)$, and the wave function plotted is the DVR state corresponding to root $\alpha=13$. The vertical phase space strip occupied by this state is marked with heavy lines.

This idea is illustrated in Fig. 5, which exhibits the phase space area occupied by the first 20 harmonic oscillator eigenstates in two different ways. On the left are plotted the first 20 annular strips, each of area $2\pi\hbar$. The Bohr–Sommerfeld quantizing orbits (not shown) lie half way between the circles in the figure, the latter of which have actions $I_n=(1/2\pi)\oint p\,dx=n\hbar$, $n=1,\ldots,20$. The wave function plotted is $\phi_{13}(x)$, which occupies the annular strip bounded by the heavy lines. On the right, the same area is divided into 20 vertical strips of equal area $2\pi\hbar$. Plotted on the x-axis are the grid points for $N=20$, which are the roots of $\phi_{20}(x)$. The wave function plotted is the DVR state $F_{13}(x)$ (roots are labeled from $\alpha=0$ at the left), which is proportional to $\phi_{20}(x)/(x-x_{13})$. It clearly vanishes at all grid points except one. The phase space area occupied by this state is the vertical strip bounded by the heavy lines.

In this way we obtain a geometrical interpretation of the unitary transformation which takes us from the orthogonal polynomial basis $\{\phi_n(x)\}$ to the DVR basis $\{F_\alpha(x)\}$. That is, it consists of dividing the phase space area occupied by the first N states into Planck cells in two different ways.

V. THE ACCURACY OF THE DVR APPROXIMATION FOR THE POTENTIAL ENERGY

We have shown that the error in the expansion of a (presumably) unknown eigenfunction in a spectral basis truncated at size N goes as e^{-cN}, when N is large enough and under appropriate conditions of analyticity and boundary conditions. Normally this implies the exponential convergence of the eigenvalues of the truncated matrix of the Hamiltonian in the spectral basis to the exact eigenvalues. If the exact energy spectrum is nondegenerate, this also implies the exponential convergence of the eigenfunctions of the truncated problem to the exact eigenfunctions, although if there are near degeneracies the basis must be large enough to resolve these (the exponentially small error must be smaller than the small energy splittings). Thus we may say, exponential convergence of the orthonormal expansion implies exponential convergence of the VBR method, to use the terminology of Ref. 12. We will not prove these statements here since their proof would take us too far astray and the issue is unrelated to the accuracy of the DVR method.

In the DVR method, however, there is an additional error due to the usual diagonal approximation for the matrix elements of the potential energy,

$$\langle F_\alpha|V|F_\beta\rangle\approx V(x_\alpha)\,\delta_{\alpha\beta}. \tag{5.1}$$

We will show below that this error is $O(1/N)$, which is much larger than the error due to the truncation of the basis alone. Nevertheless, the final error in the eigenvalues and eigenfunctions in the DVR method is still exponentially small. The discrepancy in the magnitudes of these two errors has been noted by several authors in the past, including Wei[50] and Baye *et al.*[11] In this section we shall explain this paradoxical behavior.

The diagonal approximation (5.1) applies when working in the DVR basis $\{|F_\alpha\rangle\}$ (in this section we follow the notation of Sec. III). One can also work in the basis $\{|\phi_n\rangle\}$, in practice often a spectral basis, in which case the approximation (5.1) is equivalent to a Gaussian quadrature approximation for the matrix elements of the potential energy. The computed eigenvalues are the same (thus they have the same error) since the two bases are unitarily equivalent, so for the purposes of studying the final error, either basis may be used. We have preferred to use the DVR basis because the arguments can be extended to the nonorthogonal case studied in Sec. VI.

We begin by making some comments on the analysis of Wei[50] of the accuracy of the DVR approximation. This is the most careful treatment of this question that we are aware of in the existing literature. Wei notes that the error in the diagonal approximation for the matrix elements of the potential energy in the DVR basis is actually rather large, but he makes the observation that if one transforms the DVR matrix to the spectral basis $\{|\phi_n\rangle\}$ (again in the notation of Sec. III), the (quadrature) error is concentrated at the lower right corner of the matrix, that is, it only affects matrix elements connecting states $|\phi_n\rangle$ with n near the cutoff value N. These are the states that in phase space live near the edges of the region covered by the projection operator. The analysis is particularly simple in the case of polynomial potentials, in which case (for orthogonal polynomial bases) the matrix for V is band-diagonal and one can make statements about the order of perturbation theory at which various corrections occur. Wei does not make assertions about the behavior of the error as a function of N, but he does give explicit error estimates for common cases of DVR functions and it is possible to extract from these the exponential convergence which we shall argue for below. We believe it should also be possible to use perturbation theory to derive the law of exponential convergence, at least in the case of polynomial potentials.

Consider now the matrix element $\langle F_\alpha|V|F_\beta\rangle$, in which $F_\beta(x)$ certainly belongs to $\mathcal{S}$. If it were true that $V(x)F_\beta(x)$ also belonged to $\mathcal{S}$, then Eq. (5.1) would be exact. But in fact, $V(x)F_\beta(x)$ contains components outside of $\mathcal{S}$, typically of order $1/N$. We argue this first on semiclassical grounds.

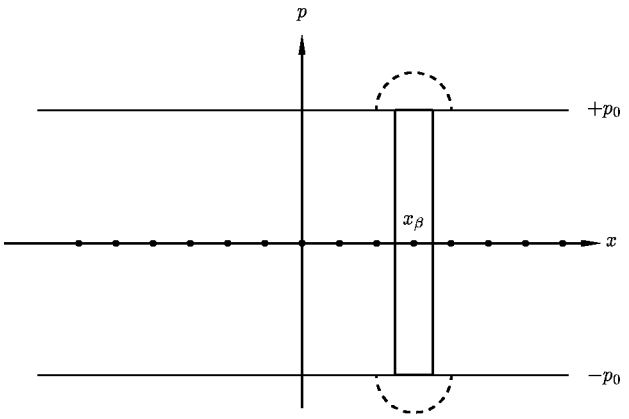

FIG. 6. When a DVR function $F_\beta(x)$ is multiplied by a slowly varying function such as $V(x)$, the resulting function is pushed a small amount beyond the bounds of the region of phase space occupied by the DVR function (at the tops and bottoms of the vertical strip, as indicated schematically by the dotted lines).

The general idea is illustrated in Fig. 6, which is a schematic illustration of the region of phase space (a vertical strip) occupied by a DVR function $F_\beta(x)$ centered on x_β. When this function is multiplied by a slowly varying function such as $V(x)$, the momentum content is modified somewhat, producing a slight ''smearing'' of the phase space region occupied by the function. This is indicated schematically by the dotted lines in the figure. In particular, the product function in general contains some phase space components which go outside the region occupied by the DVR functions (the region associated with the projection P). Any slight smearing will do this, since the DVR functions $F_\beta(x)$ extend all the way to the edge of this region.

More quantitatively, the wave number associated with the top of the momentum strip is of the order of $1/a$, where a (or Δx) is the spacing between DVR grid points. The DVR function has oscillations on this scale, or, equivalently, it has momentum components which go all the way to the top of the strip. If we let L be the scale length of the potential, then the product function $V(x)F_\beta(x)$ contains wave numbers which go like $1/a+1/L=(1/a)[1+O(a/L)]$. If $N\gg 1$ DVR functions are used to cover the range of the potential, then $a\ll L$ and $a/L=O(1/N)$. Thus we expect the relative amount by which $V(x)F_\beta(x)$ extends outside the subspace $\mathcal{S}$ to be $O(1/N)$, which implies an error of order $1/N$ in Eq. (5.1). We have examined some cases of potentials $V(x)$ for which the DVR matrix elements can be computed analytically, and confirmed the $1/N$ behavior of the error of the diagonal approximation.

To understand the exponentially small error in the DVR method, we begin with the eigenvalue problem restricted to the subspace $\mathcal{S}$, expressed in the DVR basis $\{F_\alpha(x)\}$,

$$\sum_\beta \left(\langle F_\alpha|T|F_\beta\rangle+\langle F_\alpha|V|F_\beta\rangle\right)c_\beta=Ec_\alpha, \tag{5.2}$$

where the matrix elements of T and V are computed exactly, $\{c_\alpha\}$ is the eigenvector, and $\psi(x)=\Sigma_\alpha c_\alpha F_\alpha(x)$ is the approximate eigenfunction. The only approximation here is the truncation of the basis; as argued above, this approximation introduces only exponentially small errors for eigenvalues whose eigenfunctions lie well within the region of phase space covered by P. These are the eigenfunctions we are interested in; they satisfy $P|\psi\rangle=|\psi\rangle$ plus exponentially small corrections, or $c_\alpha=\langle F_\alpha|\psi\rangle=(1/\sqrt{K_\alpha})\psi(x_\alpha)$ to exponential accuracy.

We are not allowed to set $\langle F_\alpha|V|F_\beta\rangle=V(x_\alpha)\delta_{\alpha\beta}$, since this would introduce errors of order $1/N$. Consider, however, the sum $\Sigma_\beta\langle F_\alpha|V|F_\beta\rangle c_\beta=\langle F_\alpha|V|\psi\rangle$. If $\psi(x)$ lies well within the region covered by $\mathcal{S}$, then so does $V(x)\psi(x)$, assuming $V(x)$ is analytic and slowly varying. This is because multiplying by $V(x)$ only changes the amplitude of the WKB wave function for $\psi(x)$, not its phase;[51] therefore the stationary phase points in the expansion of $V(x)\psi(x)$ in an analytic spectral basis are the same as those for the expansion of $\psi(x)$ itself. Thus we have $V|\psi\rangle=PV|\psi\rangle$ plus exponentially small corrections. The exponentially small error in the expansion of $V\psi$ will not be as good as the exponentially small error in the expansion of ψ itself, so if we are achieving convergence by increasing the size of the region covered by P, then some extra margin will be required to achieve the same accuracy in the expansion of $V\psi$ as we have in the expansion of ψ, but, assuming an analytic and slowly varying potential, both errors will be exponentially small. Thus, we can expand $V(x)\psi(x)$ in the same way we expanded $\psi(x)$ in Eq. (3.9),

$$\begin{aligned} V(x)\psi(x)&=\sum_\alpha \frac{1}{\sqrt{K_\alpha}}V(x_\alpha)\psi(x_\alpha)F_\alpha(x)\\ &=\sum_\alpha V(x_\alpha)F_\alpha(x)c_\alpha, \end{aligned} \tag{5.3}$$

plus exponentially small corrections. When this is used in Eq. (5.2), we get the usual DVR eigenvalue problem,

$$\sum_\beta \left[\langle F_\alpha|T|F_\beta\rangle+V(x_\alpha)\delta_{\alpha\beta}\right]c_\beta=Ec_\alpha. \tag{5.4}$$

Therefore the eigenvalues of the DVR method are exponentially close to the exact eigenvalues, for eigenfunctions which lie well within the region covered by P. For eigenfunctions which are near the edge of this region [these would not be well converged even in the formulation (5.2)] the approximation (5.3) is not particularly good, and their eigenvalues will be substantially changed.

This is the best argument we know of for understanding the exponential convergence which is observed in numerical experiments with DVR functions. We remark that the argument given here for exponential convergence in the DVR method does not apply to the potential-optimized DVR method.[52,53] We do not know whether the latter method gives exponential convergence.

It is interesting that the argument above can be generalized to operators other than V. For example, like V, the momentum operator p applied to $\psi(x)$ changes only the amplitude of the WKB wave function, not the phase. The same is true for any operator which is a slowly varying function of

x and p. Thus, the kinetic energy can be treated like the potential energy, so that, for eigenfunctions well within the region covered by P, we can write

$$\sum_\beta \langle F_\alpha|T|F_\beta\rangle c_\beta = \langle F_\alpha|T|\psi\rangle = \frac{1}{\sqrt{K_\alpha}}\langle x_\alpha|T|\psi\rangle$$

$$= \sum_\beta \frac{1}{\sqrt{K_\alpha}}\langle x_\alpha|T|F_\beta\rangle c_\beta, \tag{5.5}$$

plus exponentially small corrections. For example, in one dimension with $T=p^2/2$, the final matrix element is $\langle x_\alpha|T|F_\beta\rangle = (-1/2)F''_\beta(x_\alpha)$. Note that with this approximation, the matrix for the kinetic energy becomes non-Hermitian, in general. This manner of treating the kinetic energy operator has been previously discussed by Baye *et al.*[11]

VI. NONORTHONORMAL DVR BASES

In this section we consider a generalization of the DVR construction, obtained by relaxing the requirements for orthogonality. Generalizations of this type have previously been considered by Light *et al.*[18] and Szalay.[19] These differ from one another partly in the choice of the DVR-like approximation for the matrix elements of the potential energy. We make the following three contributions to this subject. First, we point out that there are two distinct bases of functions that arise in the nonorthogonal generalization of the DVR construction, one the projected δ-functions $\{\Delta_\alpha(x)\}$, and the other the interpolating basis $\{L_\alpha(x)\}$, and that these are dual to one other (they form a biorthogonal basis). This fact is independent of any DVR-like approximation for the potential energy (that is, any strategy for obviating the use of quadratures for the matrix elements of the potential energy). Second, we have pointed out that in the interpolating function basis, one can contract the basis in the usual DVR manner, that is, by throwing away grid points where the wave function is known to be small. Third, we have shown that one of Szalay's formulations of the generalized DVR method, that is, one of his DVR-like approximations for the potential energy, leads to exponential convergence. We have not, however, proposed any new approximation for the potential energy.

We begin with a modification of the presentation in Sec. III. Suppose we have a projection operator P acting on $\mathcal{H}=L^2(M)$ for some manifold M, and suppose $\{x_\alpha\}$ is a set of N grid points, where $N=\dim \mathcal{S}$ and $\mathcal{S}=P\mathcal{H}$. Suppose, however, that the DVR conditions are not satisfied, so that if $|\Delta_\alpha\rangle = P|x_\alpha\rangle$, then $\Delta_{\alpha\beta}\equiv\langle\Delta_\alpha|\Delta_\beta\rangle = \langle x_\alpha|\Delta_\beta\rangle = \Delta_\beta(x_\alpha) = \Delta_\alpha(x_\beta)^*$ is not diagonal. The matrix $\Delta_{\alpha\beta}$ is the overlap matrix of the set of projected δ-functions $\{\Delta_\alpha(x)\}$, and we may consider using this set as a nonorthogonal basis. We shall assume that this matrix is nonsingular, which means that the projected δ-functions are linearly independent.

We can still introduce a set of interpolating functions $\{L_\alpha(x)\}$, but these are not proportional to the $\{\Delta_\alpha(x)\}$, as in the DVR case. The interpolating functions are defined as the unique linear combinations of the $\{\Delta_\alpha(x)\}$ such that $L_\alpha(x_\beta)=\delta_{\alpha\beta}$. These functions are unique because of the invertibility of the matrix $\Delta_\alpha(x_\beta)$. Both sets, $\{\Delta_\alpha(x)\}$ and $\{L_\alpha(x)\}$, lie in the subspace $\mathcal{S}$, and span it, so that $P|\Delta_\alpha\rangle = |\Delta_\alpha\rangle$ and $P|L_\alpha\rangle = |L_\alpha\rangle$.

Then it turns out that the interpolating functions are dual to the projected δ-functions, that is,

$$\langle\Delta_\alpha|L_\beta\rangle = \langle L_\beta|\Delta_\alpha\rangle = \delta_{\alpha\beta}. \tag{6.1}$$

This is easily proved as follows:

$$\langle\Delta_\alpha|L_\beta\rangle = \langle x_\alpha|P|L_\beta\rangle = \langle x_\alpha|L_\beta\rangle = L_\beta(x_\alpha) = \delta_{\alpha\beta}. \tag{6.2}$$

The interpolating functions can be used in the usual way to expand arbitrary functions $\psi\in\mathcal{S}$,

$$\psi(x) = \sum_\alpha L_\alpha(x)\psi(x_\alpha). \tag{6.3}$$

In particular, consider $|\psi\rangle = P|y\rangle$, where y is fixed. Then certainly $|\psi\rangle$ belongs to $\mathcal{S}$, so the wave function $\psi(x) = \langle x|P|y\rangle$ can be expanded according to Eq. (6.3). This gives

$$\langle x|P|y\rangle = \sum_\alpha L_\alpha(x)\langle x_\alpha|P|y\rangle = \sum_\alpha \langle x|L_\alpha\rangle\langle\Delta_\alpha|y\rangle. \tag{6.4}$$

Since this is true for all x and y, we can strip off $\langle x|$ and $|y\rangle$, to obtain

$$P = P^\dagger = \sum_\alpha |L_\alpha\rangle\langle\Delta_\alpha| = \sum_\alpha |\Delta_\alpha\rangle\langle L_\alpha|. \tag{6.5}$$

Moreover, the overlap matrices $\Delta_{\alpha\beta}=\langle\Delta_\alpha|\Delta_\beta\rangle$ and $L_{\alpha\beta} = \langle L_\alpha|L_\beta\rangle$ are inverses of each other, since

$$\sum_\mu \langle L_\alpha|L_\mu\rangle\langle\Delta_\mu|\Delta_\beta\rangle = \langle L_\alpha|P|\Delta_\beta\rangle = \langle L_\alpha|\Delta_\beta\rangle = \delta_{\alpha\beta}, \tag{6.6}$$

where we have used Eqs. (6.1) and (6.5). Finally, we note that the coefficients needed to expand $L_\alpha(x)$ as a linear combination of the $\Delta_\alpha(x)$ are precisely the components of the overlap matrix $L_{\alpha\beta}$,

$$L_\alpha(x) = \sum_\beta \Delta_\beta(x)L_{\beta\alpha}, \tag{6.7}$$

as follows immediately from $L_\alpha(x_\beta)=\delta_{\alpha\beta}$ and Eq. (6.6). Thus, given the $\{\Delta_\alpha(x)\}$, we can compute $\Delta_{\alpha\beta}$ and then $L_{\alpha\beta}$ by matrix inversion, and finally the $\{L_\alpha(x)\}$.

This nonorthogonal DVR formalism presents us with three obvious choices of basis: the projected δ-functions $\{\Delta_\alpha(x)\}$, the interpolating functions $\{L_\alpha(x)\}$, and some orthonormalized version of these, $\{F_\alpha(x)\}$. Of these, Light *et al.*[18] have considered the third, while Szalay[19] has presented a more general formalism that incorporates all three. Unlike the standard DVR case, these three classes of functions are not proportional. The question is to what extent the usual advantages of DVR bases are maintained with the various choices. These advantages are the convenient approximation for the potential energy (which should not destroy the exponential accuracy in the eigenvalues and eigenfunctions) and the ability to contract the basis by throwing away points where ψ is known to be small. In the following we show that a convenient approximation for the potential energy (with exponential convergence to the final answers) is maintained

with both the projected δ-function and the interpolating function bases, and that the basis set contraction can be applied in the interpolating function basis. The potential energy approximation that leads to exponential convergence is one of the formulations of the nonorthogonal generalization of the DVR method previously considered by Szalay.

Consider first the interpolating functions as a basis for the eigenvalue problem. Let ψ be an eigenfunction that is contained well within the phase space region covered by P, so that an expansion of the form (6.3) is valid to exponential accuracy. Write this as $\psi(x)=\Sigma_\alpha c_\alpha L_\alpha(x)$, where the coefficients are given by $c_\alpha=\psi(x_\alpha)=\langle\Delta_\alpha|\psi\rangle$. Then these coefficients are the eigenvectors of

$$\sum_\beta\ (\langle L_\alpha|T|L_\beta\rangle+\langle L_\alpha|V|L_\beta\rangle-EL_{\alpha\beta})c_\beta=0. \tag{6.8}$$

Note that since c_α is the value of ψ at the grid point x_α, we can contract the basis in this formulation simply by throwing away grid points where the wave function is known to be small.

We can also make a DVR-like approximation to the matrix elements of the potential energy in Eq. (6.8). Following the steps in Sec. V and assuming that $\psi(x)$ lies well within the region covered by P, we first note that $\Sigma_\beta\langle L_\alpha|V|L_\beta\rangle c_\beta=\langle L_\alpha|V|\psi\rangle$, to exponential accuracy. Next, we note that if V is analytic and slowly varying, then $V(x)\psi(x)$ has an expansion like $\psi(x)$ in Eq. (6.3), $V(x)\psi(x)=\Sigma_\beta V(x_\beta)\psi(x_\beta)L_\beta(x)$ plus exponentially small corrections, so that $\langle L_\alpha|V|\psi\rangle=\Sigma_\beta L_{\alpha\beta}V(x_\beta)c_\beta$. Thus, the eigenvalue problem (6.8) becomes

$$\sum_\beta\ (\langle L_\alpha|T|L_\beta\rangle+L_{\alpha\beta}V(x_\beta)-EL_{\alpha\beta})c_\beta=0. \tag{6.9}$$

This is the analog of the diagonal approximation to the matrix elements of the potential energy in the standard DVR case. Note that the matrix is non-Hermitian. The eigenvalues that are well converged are real and have orthogonal eigenvectors, to exponential accuracy. The eigenvalues that are not well converged are not real, in general, nor are their eigenvectors even approximately orthogonal. It is these latter eigenvalues and eigenvectors, which we do not care about, that are responsible for the non-Hermiticity of the matrix in Eq. (6.9).

The eigenvalue problem can also be formulated in the basis $\{\Delta_\alpha(x)\}$, and again one can make DVR-like approximations on the matrix elements of the potential energy. The easiest way to derive this is to conjugate Eq. (6.9) by the matrix $\Delta_{\alpha\beta}$. This leads to

$$\sum_\beta\ (\langle\Delta_\alpha|T|\Delta_\beta\rangle+V(x_\alpha)\Delta_{\alpha\beta}-E\Delta_{\alpha\beta})d_\beta=0, \tag{6.10}$$

where the wave function is given by $\psi(x)=\Sigma_\beta d_\beta\Delta_\beta(x)$. The expansion coefficients are related to the values of the wave function by $d_\beta=\langle L_\beta|\psi\rangle=\Sigma_\gamma L_{\beta\gamma}c_\gamma$, which shows (since $c_\gamma=\psi(x_\gamma)$) that the coefficient d_β depends on the value of the wave function at all the grid points, in general. Thus, there is no easy way of contracting the basis $\{\Delta_\alpha(x)\}$.

A class of problems we have had in mind is one in which the wave function occupies perhaps only a small portion of the manifold in question. For example, scattering wave functions on the hypersphere may be highly localized, depending on the hyper-radius and the number of fragments in the final state. If hyperspherical harmonics are used as a basis,[54,55] then very large bases may be required for the short wavelengths necessary to represent the localized wave functions. In such cases the basis set contraction offered by (usual, orthonormal) DVR functions would be a great advantage, if multidimensional, orthonormal DVR functions were available on such spaces. We remark that non-DVR methods such as the hyperquantization algorithm[56] already offer such advantages, although the general framework is based on angular momentum algebra and is rather different from that presented here.

The nonorthonormal basis $\{\Delta_\alpha(x)\}$ does not allow basis set contraction in any obvious way, so it would seem that one would have to work with very large matrices (whose size is enough to cover the whole space, even though the desired wave functions are localized). Thus the exponential convergence and convenience of the potential energy matrix would not help much. In the interpolation basis $\{L_\alpha(x)\}$, however, basis set contraction is easy and would lead to much smaller matrices (of a size determined by the region of space occupied by the wave function, not by whole space). On the other hand, the functions $\{L_\alpha(x)\}$ and the overlap matrix $L_{\alpha\beta}$ themselves seem to be more difficult to determine, as they require an inversion of a large matrix (determined by the size of the space). One way out is to use a set of grid points which are the orbits of a group action on the space; for example, in the case of the ordinary 2-sphere, if points $\{x_\alpha\}$ are generated by the action of the 60-element icosahedral group, then the size of the matrices to be inverted to find $L_\alpha(x)$ or $L_{\alpha\beta}$ is reduced by a factor of 60. We shall elaborate upon these and similar considerations in future publications.

Another issue with nonorthonormal bases is that often near linear dependencies develop as the size of the basis set is increased, that is, the overlap matrix acquires very small eigenvalues. These can easily outstrip machine precision. This happens with distributed Gaussians[57] and also with overdense coherent state bases or Gabor expansions, which have been used in quantum calculations.[58] Such near linear dependencies can limit the precision with which the eigenvalues are determined, and also introduce extra parameters to be adjusted in the process of obtaining convergence. We have found, however, in numerical experiments with projected δ-function bases, that the overlap matrices can be quite stable. For example, with up to hundreds of grid points on the sphere and with P containing all spherical harmonics out to some maximum l value, the condition of the matrix $\Delta_{\alpha\beta}$ (defined as the ratio of the largest to the smallest eigenvalue) can be kept less than 10. This requires that the grid points be as equally spaced as possible; we interpret this to mean that each projected δ-function is allotted one Planck cell of phase space. We followed some variations on the schemes of Sobolev and Vaskevic[59] in the construction "equally spaced" grid points on the sphere.

We now comment on previous work on nonorthogonal

DVR bases,[18,19] which we interpret somewhat and translate into the notation of this section for comparison. We note that our overlap matrix $\Delta_{\alpha\beta}$ is denoted $\tilde{\Delta}$ in Light *et al.*,[18] while other notational differences have been summarized in Sec. III. In Light *et al.*,[18] the authors consider an orthonormalized version of the basis $\{\Delta_\alpha(x)\}$, which we define and write in our notation as

$$|F_\alpha\rangle = \sum_\beta |\Delta_\beta\rangle (\Delta^{-1/2})_{\beta\alpha}, \tag{6.11}$$

so that $\langle F_\alpha|F_\beta\rangle = \delta_{\alpha\beta}$. These authors argue that in the basis $\{F_\alpha(x)\}$ the matrix elements of the potential energy should be approximated by

$$\langle F_\alpha|V|F_\beta\rangle = V(x_\alpha)\ \delta_{\alpha\beta}, \tag{6.12}$$

which, if the DVR basis were orthogonal, would be the standard diagonal approximation. It is also noted that this approximation is exact whenever $V(x)F_\beta(x)$ belongs to the subspace $\mathcal{S}$, even for nonorthogonal DVR bases. It is, however, not the same as the diagonal-like approximation seen in Eqs. (6.9) or (6.10), which as we have shown leads to exponential convergence. We have been unable to find a reason why Eq. (6.12) should lead to exponential convergence. Szalay[19] develops another generalization of the DVR method, in fact a whole family of generalizations, which includes both that of Light *et al.* and also the formulations (6.10) and (6.9) above, but he does not seem to consider the question of basis set contraction. Szalay also presents numerical experiments testing the various methods.

VII. CONCLUSIONS

In conclusion, we have axiomatized the DVR construction by placing projection operators in a central role and integrating phase space and semiclassical concepts. We have emphasized exponential convergence and shown that it holds in the DVR method. We have discussed the standard, one-dimensional examples of DVR functions from the standpoint of our formalism. We have also discussed nonorthogonal generalizations of DVR functions, and showed that they, too, lead to exponential convergence. In future publications we will show how the framework presented in this paper can be used to develop multidimensional and other generalizations of the DVR method.

ACKNOWLEDGMENTS

The authors thank many people for useful conversations and invaluable pieces of information, especially Vincenzo Aquilanti, Joel Bowman, Simona Cavalli, Ingrid Daubechies, Don Kouri, Aron Kuppermann, Antonio Laganà, John Light, Nancy Makri, David Manolopoulos, Russell Pack, Gregory Parker, and Ned Sibert. They also thank Viktor Szalay for his careful reading of the manuscript and for pointing out several gaps in the logic. This work was supported by the National Science Foundation under Grant No. PHY-0070879 and also by the Center for Pure and Applied Mathematics, University of California, Berkeley. Tucker Carrington thanks the Alexander von Humboldt Stiftung for financial support and the theoretical chemistry group in Heidelberg for their hospitality.

[1] D. O. Harris, G. G. Engerholm, and W. D. Gwinn, J. Chem. Phys. **43**, 151 (1965).
[2] A. S. Dickinson and P. R. Certain, J. Chem. Phys. **49**, 4209 (1968).
[3] J. V. Lill, G. A. Parker, and J. C. Light, Chem. Phys. Lett. **89**, 483 (1982).
[4] Z. Bačić and J. C. Light, J. Chem. Phys. **85**, 4594 (1986).
[5] J. V. Lill, G. A. Parker, and J. C. Light, J. Chem. Phys. **85**, 900 (1986).
[6] J. C. Light and Z. Bačić, J. Chem. Phys. **87**, 4008 (1987).
[7] Z. Bačić and J. C. Light, J. Chem. Phys. **86**, 3065 (1987).
[8] Seung E. Choi and J. C. Light, J. Chem. Phys. **90**, 2593 (1989).
[9] J. Tennyson and J. R. Henderson, J. Chem. Phys. **91**, 3815 (1989).
[10] Hasan Karabulut and Edwin L. Sibert III, J. Phys. B **30**, L513 (1997).
[11] D. Baye, M. Hesse, and M. Vincke, Phys. Rev. E **65**, 026701 (2002).
[12] John C. Light and Tucker Carrington, Jr., Adv. Chem. Phys. **114**, 263 (2000).
[13] D. Baye and P.-H. Heenen, J. Phys. A **19**, 2041 (1986).
[14] O. A. Sharafeddin and J. C. Light, J. Chem. Phys. **102**, 3622 (1995).
[15] Bill Poirier, Found. Phys. **30**, 1191 (2000).
[16] Bill Poirier and J. C. Light, J. Chem. Phys. **111**, 4869 (1999).
[17] Bill Poirier and J. C. Light, J. Chem. Phys. **113**, 211 (2000).
[18] J. C. Light, I. P. Hamilton, and J. V. Lill, J. Chem. Phys. **82**, 1400 (1985).
[19] Viktor Szalay, J. Chem. Phys. **105**, 6940 (1996).
[20] Matthew Cargo and Robert G. Littlejohn, J. Chem. Phys. (submitted).
[21] M. Faist, J. Chem. Phys. **65**, 5427 (1976).
[22] François Gygi, Phys. Rev. B **48**, 11692 (1993).
[23] François Gygi, Phys. Rev. B **51**, 11190 (1995).
[24] Eyah Fattal, Roi Baer, and Ronnie Kosloff, Phys. Rev. E **53**, 1217 (1996).
[25] Bill Poirier, Phys. Rev. A **56**, 120 (1997).
[26] V. Kokoouline, O. Dulieu, R. Kosloff, and F. Masnou-Seeuws, J. Chem. Phys. **110**, 9865 (1999).
[27] Sophya Garashchuk and John C. Light, J. Chem. Phys. **114**, 3929 (2001).
[28] Matthew C. Cargo and Robert G. Littlejohn, Phys. Rev. E **65**, 026703 (2002).
[29] N. L. Balazs and B. K. Jennings, Phys. Rep. **104**, 347 (1984).
[30] M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, Phys. Rep. **106**, 121 (1984).
[31] Robert G. Littlejohn, Phys. Rep. **138**, 193 (1986).
[32] Robert G. Littlejohn, J. Math. Phys. **31**, 2952 (1990).
[33] Robert G. Littlejohn, J. Stat. Phys. **68**, 7 (1992).
[34] I. Daubechies, IEEE Trans. Inf. Theory **36**, 961 (1990).
[35] Charles K. Chui, *Wavelets: A Tutorial in Theory and Applications* (Academic, New York, 1992).
[36] T. A. Arias, Rev. Mod. Phys. **71**, 267 (1999).
[37] Kevin A. Mitchell and Robert G. Littlejohn, Phys. Rev. A **61**, 042502 (2000).
[38] M. J. Bramley, J. W. Tromp, T. Carrington, Jr., and G. C. Corey, J. Chem. Phys. **100**, 6175 (1994).
[39] V. A. Fock, Izv. Akad. Nauk SSSR, Ser. Fiz. **18**, 161 (1954); Forh. - K. Nor. Vidensk. Selsk. **31**, 138 (1958); **31**, 145 (1958).
[40] R. T. Pack, in *Advances in Molecular Vibrations and Collision Dynamics*, edited by Joel Bowman (JAI, Greenwich, 1994), Vol. 2A, p. 111.
[41] Robert G. Littlejohn, Kevin A. Mitchell, Vincenzo Aquilanti, and Simonetta Cavalli, Phys. Rev. A **58**, 3705 (1998).
[42] Robert G. Littlejohn, Kevin A. Mitchell, Matthias Reinsch, Vincenzo Aquilanti, and Simonetta Cavalli, Phys. Rev. A **58**, 3718 (1998).
[43] David Gottlieb and Steven A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications* (SIAM, Philadelphia, 1977).
[44] M. V. Berry, Proc. R. Soc. London, Ser. A **287**, 237 (1977).
[45] Kevin A. Mitchell and Robert G. Littlejohn, J. Phys. A **33**, 1 (2000).
[46] John Lund and Kenneth L. Bowers, *Sinc Methods* (SIAM, Philadelphia, 1992).
[47] Daniel T. Colbert and William H. Miller, J. Chem. Phys. **96**, 1982 (1992).
[48] Gabor Szegö, *Orthogonal Polynomials*, revised edition (American Mathematical Society, New York, 1959).
[49] Harry Hochstadt, *The Functions of Mathematical Physics* (Dover, Toronto, 1986).
[50] Hua Wei, J. Chem. Phys. **106**, 6885 (1997).
[51] In higher dimensions, the WKB wave function for ψ generally does not exist. But the Wigner function exists, and it falls off exponentially outside

the classically allowed region for the energy associated with ψ. See M. V. Berry, Proc. R. Soc. London, Ser. A **423**, 219 (1989).
[52] Julián Eschave and David C. Clary, Chem. Phys. Lett. **190**, 225 (1992).
[53] Hua Wei and Tucker Carrington, Jr., J. Chem. Phys. **97**, 3029 (1992).
[54] Aron Kuppermann, J. Phys. Chem. A **101**, 6368 (1997).
[55] Robert G. Littlejohn, Kevin A. Mitchell, and Vincenzo Aquilanti, Phys. Chem. Chem. Phys. **1**, 1259 (1999).
[56] V. Aquilanti, S. Cavalli, and D. De Fazio, J. Chem. Phys. **109**, 3792 (1998); V. Aquilanti, S. Cavalli, D. De Fazio, A. Volpi, A. Aguilar, X. Giménez, and J. M. Lucas, *ibid.* **109**, 3805 (1998).
[57] I. P. Hamilton and J. C. Light, J. Chem. Phys. **84**, 306 (1986).
[58] Michael J. Davis and Eric J. Heller, J. Chem. Phys. **71**, 3383 (1979).
[59] S. L. Sobolev and V. L. Vaskevic, *The Theory of Cubature Formulas* (Kluwer Academic, Dordrecht, Boston, 1997).