Using collocation and solutions for a sum-of-product potential to compute vibrational energy levels for general potentials

Jesse Simmons
Department of Physics, Engineering Physics and Astronomy,
Queen’s University, Kingston, ON K7L 3N6, Canada

Tucker Carrington Jr.
Department of Chemistry, Queen’s University, Kingston, ON K7L 3N6, Canada
(Dated: July 2, 2021)

Abstract
It is easier to compute a vibrational spectrum when the potential energy surface (PES) is a sum-of-products (SOP). Many popular computational methods work only if the PES is a SOP. However, the most accurate PESs are not SOPs. We propose using collocation and solutions of the Schrödinger equation with a SOP PES to compute solutions on a corresponding general PES. This makes it possible to account for coupling and anharmonicity omitted from the SOP PES. We find that correcting energy levels computed on a SOP PES with collocation reduces differences with exact energy levels by about two orders of magnitude.
I. INTRODUCTION

Given a potential energy surface (PES), it is possible to solve the time-independent Schrödinger equation numerically exactly to compute the vibrational (or ro-vibrational) spectrum of a polyatomic molecule [1–6]. Such calculations can be divided into two groups: those in which the potential energy surface (PES) has a special simplifying form and those in which the form of the PES is general. The two most common special simplifying forms are the sum-of-products (SOP) form [7–9] and the N-mode representation form [10–13]. When the PES has a special simplifying form, it is possible to do accurate calculations for molecules with tens of atoms [9, 14–16]. Calculations with general PESs are more difficult and only tractable for smaller molecules. In this Letter, we propose a method for cheaply obtaining good estimates of energies on a general PES using the results of a calculation done on a related PES with a special simplifying form. The special form we focus on is the SOP form.

The SOP form makes calculations easier because: 1) it obviates the need for full-dimensional quadrature [17]; 2) it reduces the cost the matrix-vector products (MVP) required to use an iterative eigensolver [18, 19]. To realize advantages 1) and 2), it is imperative that each basis function be a product of univariate functions of the coordinates used to represent the PES. Advantage 1) is obvious. With a SOP PES and a product basis, every Hamiltonian matrix element is a product of one-dimensional integrals. If the PES does not have SOP form then multidimensional quadrature is required. Multidimensional quadrature is costly and typically requires a large set of quadrature points. Advantage 2) is also important. If a basis has product structure it can be exploited when the PES is a SOP to efficiently evaluate MVPs by doing sums sequentially. The simplest such basis is a direct product basis and in this case the cost of a MVP scales as \( n^{D+1} \), where \( D \) is the dimensionality and \( n \) is a representative number of basis functions for a single coordinate [20]. Similar ideas can be used with a non-direct product basis [21, 22]. These and related advantages have made it possible to do calculations on large molecules [9, 14, 15].

When the PES does not have a simple form, e.g. a SOP, it becomes necessary to use (in some form) quadrature and to solve

\[
(K^{\text{exact}} + V^{\text{quad}})U = S^{\text{quad}}UE, \tag{1}
\]

where \( K^{\text{exact}} \) is an exact matrix representation of the kinetic energy operator (KEO), \( V^{\text{quad}} \) is a quadrature approximation to the matrix representation of the PES, and \( S^{\text{quad}} \) is a quadrature approximation to the overlap matrix. If the basis functions are products of univariate functions then
it is common to use a direct product quadrature. Often one chooses the 1-D quadratures so that $S_{\text{quad}} = I$. Non-direct product quadrature schemes are difficult to design [23]. The number of points in a direct product quadrature grid scales as $n^D$. If $D \gtrsim 9$ then storing the PES on the grid requires a lot of memory. A direct product discrete variable representation (DVR) calculation has similar memory requirements [20]. It is possible to use quadrature, but reduce the memory required, by dividing coordinates into two groups and solving two sub-problems. Some calculations use basis functions that are products of primitive functions for one group and eigenfunctions of one of the sub-problems [24, 25]. Others use basis functions that are products of the eigenfunctions of both sub-problems [26–28]. These calculations require tuning several parameters and using large direct product quadrature grids for the sub-problems.

In this Letter we bridge the dichotomy between the simplicity of the SOP PES world and the accuracy and generality of the general PES world. Probably the most natural way to do this is to use quadrature and basis functions for solving the Schrödinger equation with $H_{\text{gen}} = \hat{K} + V_{\text{gen}}$ that are eigenfunctions of $H_{\text{sop}} = \hat{K} + V_{\text{sop}}$. $\hat{K}$ is the kinetic energy operator (KEO); $V_{\text{gen}}$ is a general PES; and $V_{\text{sop}}$ is a SOP approximation to $V_{\text{gen}}$. These basis functions should be excellent in the sense that few will be required as long as $V_{\text{sop}} \sim V_{\text{gen}}$. How would one choose quadrature points and weights for computing matrix elements of $H_{\text{gen}}$ in the basis of the eigenfunctions of $H_{\text{sop}}$? The simplest idea is to use a direct product quadrature. It can be chosen so that $S_{\text{quad}} = I$, but a direct product grid has many points. In this letter, we shall instead use "rectangular collocation" [29–32]. Rectangular collocation (RC) has the disadvantage that it requires solving a generalized eigenvalue problem, but the advantage that it is not necessary to choose points (and there are no weights) with which quadrature approximations to integrals are accurate. We show that with RC and a point set much smaller than a direct product grid, it is straightforward to incorporate much of the effect of $V_{\text{gen}} - V_{\text{sop}}$ on the vibrational energies.

II. RECTANGULAR COLLOCATION

As is typically the case, wavefunctions $\Psi_n$ are represented as linear combinations of basis functions $\phi_i(x)$, $i = 1, \cdots, N$ that depend on coordinates $x$. In collocation[33–35], the basis expansion coefficients are determined by demanding that the Schrödinger equation be satisfied at collocation points $\{x_m\}$ for $m = 1, \cdots, M$ and solving

$$(G + VB)\tilde{U} = B\tilde{U},$$

(2)
where elements of the diagonal matrix $\tilde{E}$ are energies. In this equation, $G_{ji} = \hat{K}\phi_i(x_j)$, $V_{jk} = V(x_j)\delta_{kj}$, and $B_{ji} = \phi_i(x_j)$. To solve this equation when $M > N$, we multiply on the left by $B^T$ and obtain

$$(B^TG + B^TVB)U = (B^TB)UE. \quad (3)$$

Using more points than basis functions is expected to yield more accurate energies because using more points constrains wavefunctions at more points. Using more points also improves the condition number of $(B^TB)$.

When it is not possible to write a wavefunction as an exact linear combination of the basis functions then the corresponding energy computed from Eq. 3 will depend on the choice of the collocation points. We call error introduced by the choice of the points point-set error. The error in an eigenvalue $E_n$ is proportional [29] to $\varepsilon^2_{bn} + \varepsilon_{bn}\varepsilon_{qn}$, where

$$\varepsilon_{bn} = ||\Psi_n - \psi_{n,\text{best}}||, \quad (4)$$

with $\Psi_n$ the exact wavefunction and $\psi_{n,\text{best}}$ the best possible approximation to the wavefunction in the $\phi_i(x), i = 1, \cdots N$ basis; and $\varepsilon_{qn}$ is a measure of point-set error. Clearly, using a basis that reduces $\varepsilon_{bn}$ should reduce the effect of $\varepsilon_{qn}$. When $V_{sop} \sim V_{gen}$, it is reasonable to expect that $\varepsilon_{bn}$ will be small when the $\phi_i(x)$ are the eigenfunctions of $H_{sop}$. The number of $\phi_i(x)$ required to achieve accurate energy levels of $H_{gen} = \hat{K} + V_{gen}$ should be small. The idea we present is to use the simplicity of an SOP method as a stepping stone to solving for energies of $H_{gen}$, without needing to store large vectors or matrices. How well does this work for realistic $V_{gen}$ and $V_{sop}$? To what extent are we able to account for the effect of $V_{gen} - V_{sop}$ on energy levels? For $P_2O$ (HFCO), we find that we are able to compute fairly accurate energy levels from a basis with 150 (1000) functions.

III. CALCULATION DETAILS

We use dimensionless normal coordinates $Q_k$ and the KEO

$$\hat{K} = \sum_k \frac{\omega_k}{2} \frac{\partial^2}{\partial Q_k^2}. \quad (5)$$

The SOP method we use to generate eigenfunctions of $H_{sop}$ is the pruned basis method of Brown and Carrington [36], although in principle any SOP method could be used. The method of Brown and Carrington starts with basis of harmonic oscillator basis (HOB) functions pruned by retaining
only functions whose total HOB index is less than some threshold value. This basis is then adaptively improved by adding the nearest neighbours of the basis functions deemed most important. Those deemed important are those for which the corresponding components of the eigenvectors are large. During and after the expansion of the basis, eigenvalues and eigenvectors are computed using ARPACK [37]. Eigenvectors of $H_{sop}$ thus computed are stored. Elements of the eigenvectors are coefficients of a HOB representation of $\phi_i(Q)\ , i = 1, \cdots N$. Elements of the matrix $B$ in Eq. 2 are obtained by evaluating $\phi_i(Q)\ , i = 1, \cdots N$ at the collocation points.

There are many ways to choose collocation points. In this Letter, we report energies that we compute with points in slabs surrounding $d$-dimensional subspaces. Many authors represent a $D$−dimensional PES as a sum of $d$-dimensional terms. In chemistry, the ideas are associated with the terms "N-mode representation" and "high dimensional model representation" [10, 12]. When fitting a PES with a sum of terms each of which involves $d$ coordinates, the simplest thing to do is to use a $d$-dimensional grid for each of the terms [38]. In Ref. 39, we used related sets of points to compute energy levels with RC and a Gaussian basis. Here the same type of points are used but the basis functions are eigenfunctions of $H_{sop}$. We shall call the points N-mode representation (NMR) points. NMR points in $d$-dimensional slabs are a good choice for computing wavefunctions if coupling between different sets of $d$ coordinates is weak. For example, if an exact wavefunction is nearly a product of 1-D functions, then the wavefunction computed by putting collocation points in slabs along coordinate axes will be accurate because if the wavefunction is accurate in the slabs then all derivatives of the wavefunction are also accurate. Note that our NMR points are selected from one set of points in the full $D$−dimensional space. They are not on a grid.

We denote a set of points in a $d$−dimensional slab as NMR($d$). We start with a quasi-random multi-dimensional Sobol sequence [40] within a hypercube. Coordinate ranges are chosen to include the Gauss Hermite points associated with the primitive HOB used to compute the eigenfunctions of $H_{sop}$. We then retain points if

$$\frac{V_{\text{ceil}} - V(Q) + \Delta}{V_{\text{ceil}} + \Delta} > \text{rand} \ ,$$

where rand is a (uniformly distributed) random number in [0, 1]. We use $\Delta = 500\ cm^{-1}$ and put $V_{\text{ceil}}$ equal to at least 10,000 cm$^{-1}$ above the largest desired energy level. This set of filtered Sobol points is reduced by retaining only points that satisfy

$$\sum_{i=1}^{D} H \left( \text{abs} \left( \frac{Q_i - Q_i^{eq}}{Q_i^{\max} - Q_i^{\min}} \right) - s \right) \leq d$$

(7)
where $H$ is the Heaviside step function. The thickness of the slabs is determined by $s$.

To solve the eigenvalue problem

\[(B^T G + B^T V B)U = B^T B U E, \quad (8)\]

we do not build the $M \times N$ matrices $B$ and $G$, nor the $M \times M$ matrix $V$. Instead, we sum over points to compute

\[
(B^T B)_{ij} = \sum_{m=1}^{M} \phi_i(Q_m) \phi_j(Q_m), \\
(B^T G)_{ij} = \sum_{m=1}^{M} \phi_i(Q_m) \hat{K} \phi_j(Q_m), \\
(B^T V B)_{ij} = \sum_{m=1}^{M} \phi_i(Q_m) \phi_j(Q_m) V(Q_m), \quad (9)
\]

and store only $N \times N$ matrices. In our calculations, $N \ll M$.

**IV. RESULTS**

We have done two tests of these ideas: one with a very good $V_{sop}$ for $P_2O$ and one with a poorer $V_{sop}$ for HFCO. In both cases, collocation with the $\phi_i(x)$ basis captures much of the difference between $V_{sop}$ and $V_{gen}$. Of course, energies are more accurate if $V_{sop}$ is close to $V_{gen}$.

**A. $P_2O$**

For $P_2O$, $V_{sop}$ is the quartic SOP of Ref. 41. $V_{gen}$ is made from $V_{sop}$ by extending it so that the cubic and some quartic derivatives of $V_{gen}$ are equal to those of $V_{sop}$. To do this, we invert Taylor series for the variables $z_1 = (1 - e^{-\beta_1 Q_1^4})^{1/2}$ and $y_k = 1 - e^{-\alpha_k Q_k}$ for $k = 2, 3$ to obtain the $Q_k$ as power series in either $z_1$ or $y_k$ respectively [42]. By choosing

\[
\beta_1 = -\frac{F_{1111}}{6 \omega_1} \quad \text{and} \quad \alpha_k = -\frac{F_{kkk}}{3 \omega_k}, \quad k = 2, 3, \quad (10)
\]

we ensure that after truncating the Taylor series for $V_{gen}$ after the terms with four powers of $Q_k$, third derivatives of $V_{gen}$ with respect to $Q_k$ for $k = 2, 3$ and the fourth derivative of $V_{gen}$ with respect to $Q_1$ are exactly equal to the corresponding force constants in $V_{sop}$. 

6
The error, $\Delta^{RC} = |E^{RC}_n - E^{exact}_n|$, where $E^{RC}_n$ is an energy computed with RC and $E^{exact}_n$ is an eigenvalue of $H_{gen}$, (calculated using a direct-product ARPACK method)[37] is smaller than the SOP error, $\Delta^{sop} = |E^{sop}_n - E^{exact}_n|$, where $E^{sop}_n$ is the energy computed on the SOP PES. The mean absolute error (MAE) of the lowest 30 energies of $H_{gen}$ computed with $V_{ceil} = 14,000 \text{ cm}^{-1}$, $d = 2$, and $s = 0.1$ is shown in Fig. 1 for increasing basis size $N$ and point set size $M$. $s$ values below 0.1 work less well. $V_{gen}$ and $V_{sop}$ are very similar: the mean absolute difference (MAD) of the lowest 30 exact energies for the two potentials is about 2.72 cm$^{-1}$. $\Delta^{RC}$ is about two order of magnitude smaller than $\Delta^{sop}$.

FIG. 1: Mean absolute error of the lowest 30 collocation-corrected energies of P$_2$O as a function of basis size
B. HFCO

$V_{gen}$ for HFCO is the 12th order SOP in the supplementary material of Ref. 43. To make $V_{sop}$ we remove from $V_{gen}$ all terms whose degree is larger than four. We then remove four of the remaining terms that cause holes in our calculations. Tab. I shows a comparison between the 4th and 12th-order exact energies calculated with a DVR basis and the ARPACK eigensolver and the RC results. 1000 eigenfunctions of $V_{sop}$ and 100,000 NMR points with $d = 5$, $s = 0.1$, and $V_{ceil} = 23,000$ cm$^{-1}$ were used. For HFCO, $V_{gen}$ is much further from $V_{sop}$; the MAD for the first 25 levels is 103.00 cm$^{-1}$. Nevertheless, RC reduces energy level errors by almost two orders of magnitude. The MAEs of the first 10 and 25 collocation-corrected energies (with respect to the 12th order energies) are 1.69 and 5.15 cm$^{-1}$ respectively, both are about two orders of magnitude smaller than the corresponding MADs of exact levels computed on $V_{sop}$ and $V_{gen}$.

V. CONCLUSION

There are many methods with which it is possible to solve the Schrödinger equation to compute a vibrational spectrum that work only if the PES is a SOP. For a large molecule it is often the case that the best available PES is a SOP. For a molecule with fewer than about 6 atoms, it is now fairly straightforward to fit an accurate non-SOP PES to ab initio data. This can be done either by using physically motivated fitting functions or general machine learning methods [44, 45]. An accurate PES is rarely a SOP in the coordinates in which one wants to solve the Schrödinger equation. Solving the Schrödinger equation on a non-SOP PES requires a very different numerical approach and some form of quadrature or collocation. In this Letter we point out that by using eigenfunctions of an SOP Hamiltonian as basis functions and rectangular collocation it is possible to compute energies that are much closer to exact energies than the energies of the SOP Hamiltonian. The number of basis functions required for HFCO is about an order of magnitude larger than the number for P$_2$O. This is not due to HFCO being a larger molecule, but to the fact that the SOP PES for HFCO is much further from the accurate potential. Similar ideas could be applied to other approximate eigenfunctions, e.g. those obtained from a semi-classical calculation [46].
TABLE I: Comparison of HFCO energies. Columns 1 and 2 are exact energies computed on the 12th and 4th order PESs. Column 3 is the difference between columns 1 and 2. Column 4 contains collocation-corrected energies computed using NMR points. Column 5 is errors in the collocation-corrected energies

<table>
<thead>
<tr>
<th>4th-Order</th>
<th>12th-Order</th>
<th></th>
<th>4th-12th</th>
<th>Coll.</th>
<th></th>
<th>Coll - 12th</th>
</tr>
</thead>
<tbody>
<tr>
<td>4588.335</td>
<td>4524.684</td>
<td>63.651</td>
<td>4524.712</td>
<td>0.028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5249.686</td>
<td>5186.070</td>
<td>63.616</td>
<td>5186.861</td>
<td>0.792</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5640.448</td>
<td>5530.125</td>
<td>110.323</td>
<td>5536.911</td>
<td>6.785</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5662.681</td>
<td>5584.570</td>
<td>78.110</td>
<td>5585.601</td>
<td>1.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5910.231</td>
<td>5846.412</td>
<td>63.819</td>
<td>5847.595</td>
<td>1.184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5982.431</td>
<td>5865.012</td>
<td>117.419</td>
<td>5866.657</td>
<td>1.646</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6294.265</td>
<td>6190.265</td>
<td>104.000</td>
<td>6191.845</td>
<td>1.579</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6322.704</td>
<td>6238.159</td>
<td>84.545</td>
<td>6240.167</td>
<td>2.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6423.445</td>
<td>6358.821</td>
<td>64.624</td>
<td>6359.555</td>
<td>0.734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6569.854</td>
<td>6505.626</td>
<td>64.228</td>
<td>6506.716</td>
<td>1.090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6643.155</td>
<td>6525.396</td>
<td>117.759</td>
<td>6532.084</td>
<td>6.688</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6682.350</td>
<td>6531.084</td>
<td>151.266</td>
<td>6545.942</td>
<td>14.857</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6710.152</td>
<td>6585.799</td>
<td>124.353</td>
<td>6594.471</td>
<td>8.672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6733.084</td>
<td>6631.509</td>
<td>101.575</td>
<td>6632.959</td>
<td>1.451</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6947.240</td>
<td>6849.342</td>
<td>97.898</td>
<td>6855.046</td>
<td>5.704</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6981.857</td>
<td>6871.854</td>
<td>110.004</td>
<td>6881.059</td>
<td>9.205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7008.759</td>
<td>6890.594</td>
<td>118.166</td>
<td>6895.183</td>
<td>4.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7058.651</td>
<td>6913.830</td>
<td>144.821</td>
<td>6920.435</td>
<td>6.605</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7078.594</td>
<td>7014.638</td>
<td>63.956</td>
<td>7019.441</td>
<td>4.803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7228.427</td>
<td>7163.604</td>
<td>64.824</td>
<td>7168.595</td>
<td>4.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7302.872</td>
<td>7183.060</td>
<td>119.812</td>
<td>7189.020</td>
<td>5.960</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7329.446</td>
<td>7188.078</td>
<td>141.368</td>
<td>7199.932</td>
<td>11.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7357.967</td>
<td>7191.810</td>
<td>166.157</td>
<td>7205.534</td>
<td>13.724</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7362.584</td>
<td>7238.030</td>
<td>124.555</td>
<td>7245.467</td>
<td>7.437</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7391.750</td>
<td>7277.498</td>
<td>114.252</td>
<td>7282.879</td>
<td>5.382</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgments

This work has been supported by the Natural Sciences and Engineering Research Council of Canada. We thank Ove Christiansen for sending us the HFCO PES.


