Locally Optimized Coordinates in Modified Shepard Interpolation[†]

Christian R. Evenhuis and Michael A. Collins*

Research School of Chemistry, Australian National University

Received: November 25, 2008; Revised Manuscript Received: February 4, 2009

An extension of the modified Shepard interpolation method is presented that allows expansions for the potential energy using different local coordinate sets to be used in a global interpolation. The coordinates used in a given Taylor expansion are determined using a training set of geometries at which the ab initio potential energy is known and that is built up during the construction of the interpolated potential energy surface. The method is applied to the bound state potential energy surface of methanol and a significant improvement in the rate of convergence of the interpolated potential energy surface to the ab initio potential energy is observed.

Introduction

To simulate the reaction dynamics of a molecular system, or to calculate the vibrational spectra of a molecule, the value of the potential energy needs to be known for large numbers of molecular geometries. The potential energy of a molecule can be calculated using modern quantum chemical methods and it is possible to employ direct or "on-the-fly" dynamics. However, as the computational cost of ab initio calculations rises steeply with the accuracy obtained, and with the number of electrons, a compromise must be made between the cost and the quality of the potential surface. As the potential needs to be known repeatedly at similar geometries in a dynamics calculation, it is more efficient to use an approximate potential energy surface (PES) that is generated by some fitting or interpolation approach. The fitting or interpolation error, the difference between the ab initio potential energy and the interpolated potential surface, introduces another source of uncertainty to the dynamics calculation.

In a traditional fitting approach,[1,2] the potential energy is expressed as a combination of functional forms, and the details of the combination are determined by minimizing the fitting error over a set of geometries at which the ab initio energy is known (a training set). Examples of fitted potential surfaces include refs 3−7. More automated approaches to potential surface generation include the reproducing Kernal Hilbert space,[8−10] neural networks,[11−15] interpolating moving least squares,[16−18] and modified Shepard interpolation.[19−22]

The latter method has a number of attractive features: the approach involves only a handful of fixed parameters, the method is local in nature, and the quality of the interpolated potential surface can be easily and systematically improved by adding more data. As a result, modified Shepard interpolation can be set up to require minimal user input and can be thought of as a "black box" approach for generating potential surfaces.

Bound state PESs are needed for vibrational calculations, for simulation of clusters, and in reactions of large molecules that are treated using a fragmentation approach.[23] In a fragmentation approach the PES for a large molecule is constructed from PESs for medium sized fragments and from potential energy correction surfaces (PECS). As PECS have been shown to be relatively easy to interpolate, there is evidence[24] that the potential surfaces of medium sized molecular fragments may be the bottleneck for producing accurate interpolated PESs for large molecules.

In this paper we present a number of refinements that aim to improve the convergence of interpolated PESs for medium sized bound state molecules. Building on previous work in which alternatives to the interpolation coordinates were investigated,[25] we allow a wider range of coordinates to be used and, more importantly, outline how different sets of local coordinates can be employed in a global interpolation scheme. Second, we outline an automated approach for selecting these local coordinates. These modifications have been applied to the bound state PES for methanol and a significant improvement in the convergence of the interpolation errors has been observed.

Theory

Many research groups have applied modified Shepard interpolation to potential surface generation, with each taking a different approach (for example see refs 21, 22, 26−32). In this section we briefly outline the interpolation scheme developed by Collins and co-workers,[19,20,33−35] discuss the use of other coordinates, and demonstrate how it is possible to use a variety of coordinate systems.

Modified Shepard Interpolation. In the modified Shepard scheme, the PES is expressed as a weighted sum of Taylor expansions for the potential energy from a scattered set of molecular geometries called data points, $\{\mathbf{X}(n)\}_{n=1}^{n_{\text{data}}}$, where, for a molecule containing n_{atom} atoms, $\mathbf{X}$ is a $(3 \times n_{\text{atom}})$-dimensional vector of Cartesian coordinates. In general, the interpolation coordinates are an overcomplete set of curvilinear coordinates, $Z = \{Z_l(X)\}_{l=1}^{n_{\text{var}}}$. Using these coordinates, we can estimate the potential energy at a given geometry as a weighted sum of second-order Taylor series expansions from the set of data points:

$$V^{\text{int}}(\mathbf{Z}) = \sum_{n=1}^{n_{\text{data}}} w(\mathbf{Z};n)\, T(\mathbf{Z};n) \qquad (1)$$

where $T(\mathbf{Z};n)$ is a second-order Taylor series expansion of the PES near $Z[X(n)]$ and $w(\mathbf{Z};n)$ is a weight function. The approach of Collins et al. actually performs the interpolation using the inverse of the interatomic distances as the basic coordinates. If

[†] Part of the "George C. Schatz Festschrift".

10.1021/jp8103722 CCC: $40.75 © 2009 American Chemical Society
Published on Web 03/13/2009

there are more than four atoms, this is an overcomplete or redundant set of coordinates.

The weights can be viewed as a probabilistic estimate of the accuracy of the Taylor series. This naturally leads to a number of conditions placed on the weights, such as the constraint that the weights must sum to unity. A simple form for the weight function that satisfies these requirements is given by

$$w(\mathbf{Z};n) = \frac{v(Z;n)}{\sum_{m=1}^{n_{\text{data}}} v(\mathbf{Z};m)} \tag{2}$$

in which $v(\mathbf{Z};n)$ is either the "one-part weight function",

$$v(\mathbf{Z};n) = \frac{1}{\left(\sum_{l=1}^{n_{\text{var}}} [Z_l - Z_l(n)]^2\right)^p} \tag{3}$$

or the "two-part weight function",[35]

$$v(\mathbf{Z};n) = \frac{1}{\left(\sum_{l=1}^{n_{\text{var}}} \left[\frac{Z_l - Z_l(n)}{d_l(n)}\right]^2\right)^p + \left(\sum_{l=1}^{n_{\text{var}}} \left[\frac{Z_l - Z_l(n)}{d_l(n)}\right]^2\right)^q} \tag{4}$$

in which $q = 2$ and $2p > 3n_{\text{atom}} - 3$. The two-part weight function associates a set of confidence lengths, $d_l(n)$, with each data point. The confidence lengths are calculated from a Bayesian analysis of the errors in the Taylor series from the nth data point to neighboring data points and can be viewed as estimates of the accuracy of the Taylor expansions in each coordinate direction. This leads to two limiting behaviours; at short-range the weight decays in proportion to the error in the Taylor expansion and at long-range the weight decays in proportion to the dimension of the system. The confidence lengths are defined as

$$d_l(n)^{-6} = \frac{1}{ME_{\text{tol}}{}^2} \sum_{m=1}^{M} \frac{\left[\frac{\partial V}{\partial Z_l}\Big|_{Z(m)} - \frac{\partial T(\mathbf{Z};n)}{\partial Z_l}\right]^2 [Z_l(m) - Z_l(n)]^2}{\|\mathbf{Z}(m) - \mathbf{Z}(n)\|^6} \tag{5}$$

where M is the number of neighboring data points and E_{tol} is the error in the Taylor series that determines the change from short to long-range behavior. The use of the two-part weight function has been shown to significantly improve interpolation accuracy.

The Taylor expansion for the potential energy about the nth data point is

$$T(Z;n) = V\Big|_{\mathbf{Z}(n)} + \sum_{l=1}^{n_{\text{var}}} \frac{\partial V}{\partial Z_l}\Big|_{\mathbf{Z}(n)} [Z_l - Z_l(n)] + \frac{1}{2} \sum_{l,k=1}^{n_{\text{var}}} \frac{\partial^2 V}{\partial Z_l \partial Z_k}\Big|_{\mathbf{Z}(n)} [Z_l - Z_l(n)][Z_k - Z_k(n)] \tag{6}$$

The expansion coefficients are calculated from ab initio calculations of the potential energy, gradient, and Hessian. Note that this ab initio data are also valid at any geometry that is related to $\mathbf{X}(n)$ by permutation of any indistinguishable nuclei, or by inversion of the Cartesian coordinates, simply by permuting/inverting the gradient and Hessian data. Hence, it is an easy matter to incorporate the correct symmetry in the PES of eq 1, simply by adding all these permuted-inverted configurations to the data set included in eq 1.

It remains to relate the ab initio Cartesian derivatives to the Taylor expansion coefficients. At a given point, the change in Cartesian coordinates are related to the change in the interpolation coordinates through the Jacobian, $\mathbf{B}$, of the mapping from Cartesians to interpolation coordinates,

$$\delta Z_l = \sum_{i=1}^{3n_{\text{atom}}} \frac{\partial Z_l}{\partial X_i} \delta X_i \equiv \sum_{i=1}^{3n_{\text{atom}}} B_{l,i} \delta X_i \tag{7}$$

If $\mathbf{B}$ can be inverted,

$$\frac{\partial X_i}{\partial Z_l} \equiv B_{l,i}{}^{-1} \text{ such that } \frac{\partial X_i}{\partial Z_l}\frac{\partial Z_l}{\partial X_i} = \delta_{i,j} \tag{8}$$

the derivatives of the potential energy in the two coordinate systems can be related,

$$\begin{aligned} \frac{\partial V}{\partial X_i} &= \sum_{l=1}^{n_{\text{var}}} \frac{\partial Z_l}{\partial X_i} \frac{\partial V}{\partial Z_l} \\ \Rightarrow \frac{\partial V}{\partial Z_l} &= \sum_{i=1}^{3n_{\text{atom}}} \frac{\partial X_i}{\partial Z_l} \frac{\partial V}{\partial X_i} \end{aligned} \tag{9}$$

Given $\mathbf{B}^{-1}$, the first-order Taylor series coefficients are given by eq 9, and by application of the chain rule the second-order Taylor series coefficients are given by

$$\begin{aligned} \frac{\partial^2 V}{\partial X_i \partial X_j} &= \sum_{k=1}^{n_{\text{var}}} \left(\frac{\partial V}{\partial Z_k} \frac{\partial^2 Z_k}{\partial X_i \partial X_j} + \sum_{l=1}^{n_{\text{var}}} \frac{\partial V}{\partial Z_l \partial Z_k} \frac{\partial Z_k}{\partial X_j} \frac{\partial Z_l}{\partial X_i} \right) \\ \Rightarrow \frac{\partial^2 V}{\partial Z_l \partial Z_m} &= \sum_{i,j=1}^{3n_{\text{atom}}} \left(\frac{\partial^2 V}{\partial X_i \partial X_j} - \sum_{k}^{n_{\text{var}}} \frac{\partial V}{\partial Z_k} \frac{\partial^2 Z_k}{\partial X_i \partial X_j} \right) \frac{\partial X_i}{\partial Z_l} \frac{\partial X_j}{\partial Z_m} \end{aligned} \tag{10}$$

However, in general, $\mathbf{Z}$ is an overcomplete set of coordinates and there are many possible solutions to eq 9; as there are n_{var} unknown Taylor coefficients and only $n_{\text{int}} = (3n_{\text{atom}} - 6)$ pieces of information in the Cartesian gradient vector, the solutions form an $n_{\text{var}} - n_{\text{int}}$ dimensional space. The singular value decomposition[36] of $\mathbf{B}$ allows us to construct the pseudoinverse, $\mathbf{B}^{-1}$, that gives the smallest solution in the least-squares sense. The matrix $\mathbf{B}$ can be expressed as

$$B_{1,i} = \sum_{k=1}^{n_{\text{var}}} \sum_{j=1}^{3n_{\text{atom}}} U_{l,k} \Lambda_{k,i} V_{j,i} \tag{11}$$

where U is an $n_{\text{var}} \times n_{\text{var}}$ orthogonal matrix, V is a $(3n_{\text{atom}}) \times (3n_{\text{atom}})$ orthogonal matrix, and Λ is an $n_{\text{var}} \times (3n_{\text{atom}})$ matrix whose only nonzero elements are its diagonal entries, which are known as the singular values, $\{\lambda_l\}_{l=1}^{(3n_{\text{atom}})}$. As the expansion coordinates are invariant under rotation and translation, six of the singular values are zero, leaving only $n_{\text{int}} = (3 \times n_{\text{atom}} - 6)$ singular values that are nonzero. Only the rows of U and V corresponding to the n_{int} nonzero singular values are used to construct the pseudoinverse,

$$B_{i,l}^{-1} = \sum_{\alpha=1}^{n_{\text{int}}} V_{i,\alpha}\lambda_{\alpha}^{-1}U_{l,\alpha} \tag{12}$$

This is the least-squares solution of eq 9; other solutions include contributions from rows of U that correspond to the null space of $\mathbf{B}$. Other inverses can be constructed by including more of the rows of U. As the rows of U are orthonormal, any contributions to in $\mathbf{B}^{-1}$ from rows of U for $\alpha > n_{\text{int}}$ will increase the matrix norm of $\mathbf{B}^{-1}$. Without further information, the least-squares solution for $\mathbf{B}^{-1}$ is the most reasonable choice to take.

Alternate Interpolation Coordinates. The inverse bond lengths were originally used as interpolation coordinates for two reasons: the bond lengths are the smallest set of invariants that can give a global description of the molecular structure,[37] and their inverses were found to result in more accurate Taylor expansions for bond stretching potential functions. Coordinates other than the inverse bond lengths have been employed in the modified Shepard interpolation: for example, dot-cross coordinates,[38] valence coordinates,[28] Cartesian normal modes,[29] and functions of the valence coordinates.[26,39] Not surprisingly, it has been found that using physically motivated functional forms as interpolation coordinates lead to more accurate Taylor expansions.

The first modification to the modified Shepard interpolation scheme considered here is to increase the set of interpolation coordinates beyond the inverse bond lengths. This larger set of coordinates is denoted $\mathbf{Z}^s$. To describe a molecule that undergoes limited displacements about some equilibrium geometry, it is not necessary to use a global set of coordinates. In this case choosing a set of interpolation coordinates is no harder than designing an appropriate $\mathbf{Z}$-matrix. In previous work that used functions of valence coordinates in Modified Shepard interpolation,[21,25,26] the number of interpolation coordinates have been equal to, or only slightly greater than, the minimal number, $(3n_{\text{atom}} - 6)$. We consider the case in which the number of expansion coordinates can be much greater than $(3n_{\text{atom}} - 6)$. An example for which a large number of coordinates would be necessary is a large molecule undergoing a structural rearrangement in which the bonding pattern changes, for example, an isomerization that occurs by intramolecular hydrogen transfer. Previously it has been suggested to use inverse bond lengths for stretching coordinates, the cosine of the bond angle for bond bending motions, and the sine and cosine of the dihedral angle for torsional motions.[39] If all possible bond lengths, bond angles and dihedrals are used, the total number of coordinates is proportional to the fourth power of n_{atom}.

Although it is possible to simply include all possible coordinates, this has a number of undesirable consequences. First, if there are a large number of interpolation coordinates, the evaluation of a Taylor series becomes much more costly. A direct expansion using n_{var} coordinates requires $n_{\text{var}}(n_{\text{var}} + 1)/2$ operations, which means that if n_{var} is proportional to the fourth power of n_{atom}, the computational cost of a single expansion increases as the eighth power of n_{atom}. In the approach of Collins et al. this cost is reduced by first transforming from the interpolation coordinates to the n_{int} "local internal coordinates" in which the Taylor series coefficients are diagonal.[34] In this approach, the number of operations to evaluate a single Taylor series is $n_{\text{var}} \times n_{\text{int}} + 2n_{\text{int}}$ the first term arises from the transformation to the local internal coordinates and the second from the Taylor expansion in local internal coordinates. If n_{var} is proportional to the fourth power of n_{atom} then the computational cost of a Taylor series increases as the fifth power of n_{atom}. Although this is significantly better scaling, it remains that the computational cost of interpolation rises steeply with the size of the molecule.

The second undesirable consequence of including every coordinate is that, for a given arrangement of the molecule, most of the coordinates are not physically relevant and one would expect coefficients of extraneous coordinates to be vanishingly small. However, the least-squares solution to eq 9 will tend to produce expansion coefficients that are all of the same order of magnitude.

The third disadvantage of using a large number of interpolation coordinates is that there are geometries at which the dihedral angle is undefined; the dihedral is undefined if three of the atoms involved in the definition of a dihedral angle are collinear. Near such a collinearity using an expansion coordinate based solely on the dihedral angle will cause the inversion of $\mathbf{B}$ to be numerically unstable. In previous work this problem has been avoided by ensuring that all of the expansion coordinates are well behaved for all geometries.[38] In this example, one could multiply the dihedral angle by the sine of the bending angles between the atoms. A more elegant solution is to exclude any problematic coordinates from the expansion.

Modification of the Interpolation Scheme. The obvious solution to the problems outlined above is to include a wide variety of coordinates in $\mathbf{Z}^s$ so that many different arrangements of the molecule can be described but at each data point to use only a small subset to perform the Taylor expansion. This is consistent with the local nature of modified Shepard interpolation, in that, as long as the weights in the interpolation formula enforce locality, it is irrelevant how the local estimates are obtained. The interpolation formula is modified to

$$V^{\text{int}}(\mathbf{Z}) = \sum_{n=1}^{n_{\text{data}}} w(\mathbf{Z}^{w};n)\, T(\mathbf{Z}^{n};n) \tag{13}$$

where $\mathbf{Z}^w$ is the set of weight function coordinates and $\mathbf{Z}^n \subset \mathbf{Z}^s$ is the set of expansion coordinates specific to the nth data point. Ideally, the number of coordinates in $\mathbf{Z}^n$ should be of the order of n_{int}, e.g., the n_{int} valence coordinates used in a $\mathbf{Z}$-matrix and perhaps a small number of extra coordinates, to allow redundancy. At the nth data point the Taylor expansion for the potential energy is performed using the local coordinates, $\mathbf{Z}^n \subset \mathbf{Z}^s$,

$$T(\mathbf{Z}^n;n) = V\Bigg|_{\mathbf{Z}^n(n)} + \sum_{Z_l \in \mathbf{Z}^n} \frac{\partial V}{\partial Z_l}\Bigg|_{\mathbf{Z}(n)} [Z_l - Z_l(n)] + \frac{1}{2}\sum_{Z_1,Z_k \in \mathbf{Z}^n} \frac{\partial^2 V}{\partial Z_l \partial Z_k}\Big|_{\mathbf{Z}(n)} [Z_l - Z_l(n)][Z_k - Z_k(n)] \tag{14}$$

The Taylor expansion coefficients are calculated as described in the previous section (eqs 9and 10), the only change being that the coordinates used are restricted to the local coordinates $\mathbf{Z}^n$.

Using the set of coordinates $\mathbf{Z}^s$ to calculate the weight functions is also undesirable. Not only is it computationally more expensive to do so, but also the measure of distance can become distorted; as the number of bonds, angles, and dihedrals increases differently with the size of the molecule, the simple Euclidean distance will overemphasize certain types of coordinates. So a smaller set of weight function coordinates, $\mathbf{Z}^w \subset \mathbf{Z}^s$, are chosen to be used in the weight function. In principle, as with the expansion coordinates, each data point could have its own set of weight coordinates. However, care would need to be taken to ensure that the measure of distance from different data points is consistent. To avoid this complication, in this work we consider a global weight function coordinate set.

As the weight function is used only to measure the distance between geometries, it may be possible to use a set of coordinates simpler than those used for the Taylor expansions. The obvious choice is the inverse distances, as they are the smallest set of invariants that can globally describe the PES. If only a small region of the PES needs to be described, for example, a molecule undergoing small amplitude molecular motions, the number of global weight coordinates needs to be only of the order of n_{int}. Given an appropriate set of weight coordinates, $\mathbf{Z}^w$, the expression for the weight function from eqs 2−4 are modified to use only $\mathbf{Z}^w$. Similarly, eq 5 is modified to calculate the confidence lengths; however, this expression is more complicated as it is necessary to evaluate the derivative with respect to weight function coordinates of the Taylor series from the nth data point,

$$d_l(n)^{-6} = \frac{1}{ME_{\text{tol}}{}^2}\sum_{m=1}^{M}\frac{\left[\left.\frac{\partial V}{\partial Z_l^w}\right|_{\mathbf{Z}(m)} - \frac{\partial T(\mathbf{Z};n)}{\partial Z_l^w}|_{\mathbf{Z}(m)}\right]^2[Z_l^w(m) - Z_l^w(n)]^2}{\|\mathbf{Z}^w(m) - \mathbf{Z}^w(n)\|^6} \tag{15}$$

The derivative of the nth Taylor series at the mth data point is more complicated to evaluate as the Taylor series is expanded in the coordinate system of the nth data point, $\mathbf{Z}^n$. The internal coordinate derivatives of the nth Taylor series are first converted to Cartesian derivatives, then converted to derivatives with respect to the weight function coordinates in the same way outlined in the previous section,

$$\left.\frac{\partial T(\mathbf{Z};n)}{\partial Z_l^w}\right|_{\mathbf{Z}(m)} = \sum_{i=1}^{3n_{\text{atom}}}\sum_{l\in\mathbf{Z}^n}\left.\frac{\partial X_i}{\partial Z_l^w}\right|_{\mathbf{Z}(m)}\left.\frac{\partial Z_k^n}{\partial X_i}\right|_{\mathbf{Z}(m)}\frac{\partial T(\mathbf{Z};n)}{\partial Z_k^n}|_{\mathbf{Z}(m)} \tag{16}$$

As the derivatives of the Taylor series with respect to the weight function coordinates are determined by the coordinates in $\mathbf{Z}^n$, the confidence lengths for the nth data point depend on which coordinates were chosen for $\mathbf{Z}^n$.

Relation between Coordinate Systems. So far we have discussed a number of coordinate systems ($\mathbf{Z}$, $\mathbf{Z}^s$, $\mathbf{Z}^w$, and $\mathbf{Z}^n$) that could be used to expand the potential energy surface. Although a different choice of coordinates will result in different coefficients in the Taylor expansion, the first and second Cartesian derivatives of the expansion will all be equivalent; where the expansions differ is in their higher-order Cartesian derivatives. It is well-known that the expansion of a PES is more compact if a well-chosen set of coordinates is used. The first reason for this is that the Taylor expansion for the potential along a single coordinate will converge more rapidly, so fewer terms are needed for an accurate description. Second, the potential surface may be nearly separable in a well-chosen set of coordinates, so only a small number of terms coupling different coordinates are needed. In other words, in a well-chosen coordinate system the higher-order diagonal derivatives are small and the mixed derivatives even smaller. If the higher-order derivatives are smaller, then a second-order Taylor expansion will be more accurate. We aim to use a good choice of expansion coordinates at each data point and so describe more of the higher order features of the potential.

Selecting Coordinate Systems. One way to select local coordinates would be to explicitly specify which coordinates to use at each data point. A more desirable approach would be to use some algorithm that, given a Cartesian geometry, selects a reasonable set of coordinates.[41−43] However, such an approach relies on complicated algorithms that use a large amount of chemical knowledge about the different types of molecular bonding. In addition, much of this chemical knowledge comes from equilibrium or transition state structures of well characterized molecules and so may be not be robust when applied to systems with large excess energy or to novel molecules.

The approach that we have developed to select the local coordinate sets is reminiscent of a traditional fitting approach in which the coefficients in an expansion for the potential are determined by minimizing the error over a training set. However, we use a training set to select which subset of coordinates to use (for more detail see Appendix II). In short, the local coordinates are chosen by steepest descent minimization of the interpolation error over a set of nearby geometries at which the ab initio energy is known. In this case, the training set we use are the candidate geometries that were accumulated during iterative refinement, and reweighted by a Metropolis sample as described in section IIE.

To ensure a good choice of coordinate system is made, several starting guesses are used, usually corresponding to different possible bonding schemes.

Derivatives. For an efficient molecular dynamics simulation it is necessary to have an analytic expression for the derivative of the interpolated potential energy. The derivative of the interpolated energy from eq 13 is

$$\frac{\partial V^{\text{int}}}{\partial X_i} = \frac{\partial V^{\text{int}}}{\partial Z_l^s}\frac{\partial Z_l^s}{\partial X_i} = \left\{\sum_{l=1}^{n_{\text{data}}}\frac{\partial}{\partial Z_l^s}w[\mathbf{Z}^w(\mathbf{X});n]T[\mathbf{Z}^n(\mathbf{X});n] + w[\mathbf{Z}^w(\mathbf{X});n]\frac{\partial}{\partial Z_l^s}T[\mathbf{Z}^n(\mathbf{X});n]\right\}\frac{\partial Z_l^s}{\partial X_i} \tag{17}$$

As $\mathbf{Z}^n \subset \mathbf{Z}^s$ the derivative of the Taylor expansion in the above is simply

$$\frac{\partial}{\partial Z_l^s}T[\mathbf{Z}^n(\mathbf{X});n] = \begin{cases} 0 & \text{if } Z_l^s \notin \mathbf{Z}^n \\ \frac{\partial}{\partial Z_k^n}T[\mathbf{Z}^n(\mathbf{X});n] & \text{if } Z_k^n Z_l^s \end{cases} \tag{18}$$

A similar expression holds for the derivative of the weight function. As the Cartesian derivatives of $\mathbf{Z}^s$ are calculated once, the computational cost of evaluating the derivative of the interpolated potential energy is only marginally more expensive than evaluating the interpolated energy.

TABLE 1: Definition of Coordinate System I (Details in Text)

Weight Function and Expansion Coordinates				
$R_{1,2}{}^{-1}$	$R_{1,3}{}^{-1}$	$R_{1,4}{}^{-1}$	$R_{1,5}{}^{-1}$	$R_{1,6}{}^{-1}$
$R_{2,3}{}^{-1}$	$R_{2,4}{}^{-1}$	$R_{2,5}{}^{-1}$	$R_{2,6}{}^{-1}$	$R_{3,4}{}^{-1}$
$R_{3,5}{}^{-1}$	$R_{3,6}{}^{-1}$	$R_{4,5}{}^{-1}$	$R_{4,6}{}^{-1}$	$R_{5,6}{}^{-1}$

Iterative Refinement. In this work we have used an iterative refinement scheme similar to that of Moyano and Collins.[40] During a single iteration, Metropolis sampling of the interpolated PES generates a set of 1000 geometries. From this set, four geometries are selected; the one with the largest *h*-weight,[19] the one with the largest variance[33] and two randomly selected geometries. The ab initio potential energy of these four points is calculated, and their geometries and energies added to a list of "candidate" geometries. The configuration that has the largest difference between the exact energy and the energy calculated using the current data set is chosen as the next data point. As the ab initio energy is known at the accumulated candidate geometries, this set provides a direct probe of the ab initio surface and is analogous to a training set that one would use in a fitting approach.

Results

Computational Details. To test the proposed interpolation scheme, interpolated potential energy surfaces were constructed for methanol at the Hartree−Fock level of theory using Dunning's augmented correlation consistent double-ζ basis. As in other studies, although an inexpensive level of ab initio theory has been employed to allow the convergence properties to be thoroughly investigated, the conclusions are transferable to higher levels of ab initio theory. The first data point was placed at the equilibrium structure of methanol and the interpolated potential energy surface was grown using the iterative refinement procedure described above. The PES treats all three methyl hydrogens equivalently, by adding six configurations that are permutations of these three hydrogens, and inversion of the whole molecule, to the data set at each iteration. However, when referring to the number of data points that define the PES, only one version of the six is taken into account. The potential surface was sampled using a modified version of Metropolis sampling (see Appendix 1) that is designed to ensure that the torsional mode is efficiently sampled and that the energy distribution in the torsional mode is displaced to much lower energies than the distribution for the higher frequency bending and stretching modes. This is achieved by separately sampling the $-CH_3$ and $-OH$ modes from the CH_3OH torsional mode. For the $-CH_3$, $-OH$, and the CH_3OH torsional modes, $\beta = 1/kT$ values of 150, 100, and 100 were used.

Choice of Coordinates. In this section we examine how the choice of coordinate systems affects the convergence of the interpolated PESs. The selection of expansion coordinates is guided by the knowledge of the likely forms of the molecular potential energy functions.

The three coordinate systems used, coordinates systems I−III, are given in Tables 1−3. The ordering of the atoms is the first atom is C, the O atom is the second atom, the methyl hydrogens are atoms 3−5, and the hydroxyl hydrogen is atom 6. The bond lengths, valence bending coordinate and dihedrals are denoted as R, θ, and ϕ. Using these conventions, $R_{1,2}$ is the C−O bond length, $\theta_{1,2,6}$ is the C−O−H bending angle, and $\phi_{3,1,2,6}$ is a torsional coordinate.

Coordinate System I. The inverse bond lengths are the simplest coordinate system that we will consider and is a benchmark for the standard modified Shepard interpolation approach. The amount of chemical intuition used in this coordinate system is minimal as one only assumes that the potential energy depends on the interatom distances and that the inverse distance is a suitable form for the potential. As a result, this coordinate system is completely general and can be applied to any molecular system.

TABLE 2: Definition of Coordinate System II (Details in Text)

Weight Function Coordinates				
$R_{1,2}{}^{-1}$	$R_{1,3}{}^{-1}$	$R_{1,4}{}^{-1}$	$R_{1,5}{}^{-1}$	$R_{1,6}{}^{-1}$
$R_{2,3}{}^{-1}$	$R_{2,4}{}^{-1}$	$R_{2,5}{}^{-1}$	$R_{2,6}{}^{-1}$	$R_{3,4}{}^{-1}$
$R_{3,5}{}^{-1}$	$R_{3,6}{}^{-1}$	$R_{4,5}{}^{-1}$	$R_{4,6}{}^{-1}$	$R_{5,6}{}^{-1}$
Expansion Coordinates				
$R_{1,2}{}^{-1}$	$R_{1,3}{}^{-1}$	$R_{1,4}{}^{-1}$	$R_{1,5}{}^{-1}$	$R_{2,6}{}^{-1}$
$\cos\theta_{2,1,3}$	$\cos\theta_{2,1,4}$	$\cos\theta_{2,1,5}$	$\cos\theta_{1,2,6}$	
$\cos\theta_{3,1,4}$	$\cos\theta_{3,1,5}$	$\cos\theta_{4,1,5}$		
$\cos\phi_{3,1,2,4}$	$\cos\phi_{3,1,2,5}$	$\cos\phi_{3,1,2,6}$	$\cos\phi_{2,1,3,4}$	$\cos\phi_{2,1,3,5}$
$\cos\phi_{4,1,2,5}$	$\cos\phi_{4,1,2,6}$	$\cos\phi_{2,1,4,3}$	$\cos\phi_{2,1,4,5}$	$\cos\phi_{5,1,2,6}$
$\cos\phi_{2,1,5,3}$	$\cos\phi_{2,1,5,4}$	$\cos\phi_{4,1,3,5}$	$\cos\phi_{3,1,4,5}$	$\cos\phi_{3,1,5,4}$

TABLE 3: Definition of Coordinate System III (Details in Text)

Weight Function Coordinates				
$2R_{1,3}{}^{-1}$	$2R_{1,4}{}^{-1}$	$2R_{1,5}{}^{-1}$	$2R_{1,2}{}^{-1}$	$2R_{2,6}{}^{-1}$
$^1/_2\cos\theta_{2,1,3}$	$^1/_2\cos\theta_{2,1,4}$	$^1/_2\cos\theta_{2,1,5}$	$^1/_2\cos\theta_{1,2,6}$	
$^1/_2\cos\theta_{3,1,4}$	$^1/_2\cos\theta_{3,1,5}$	$^1/_2\cos\theta_{4,1,5}$		
$^1/_4\cos\phi_{6,2,1,3}$	$^1/_4\cos\phi_{6,2,1,4}$	$^1/_4\cos\phi_{6,2,1,5}$		
Expansion Coordinates				
$2R_{1,3}{}^{-1}$	$2R_{1,4}{}^{-1}$	$2R_{1,5}{}^{-1}$	$2R_{1,2}{}^{-1}$	$2R_{2,6}{}^{-1}$
$2\exp(-R_{1,3})$	$2\exp(-R_{1,4})$	$2\exp(-R_{1,5})$	$2\exp(-R_{1,2})$	$2\exp(-R_{2,6})$
$^1/_2\cos\theta_{2,1,3}$	$^1/_2\cos\theta_{2,1,4}$	$^1/_2\cos\theta_{2,1,5}$	$^1/_2\cos\theta_{1,2,6}$	
$^1/_2\sin\theta_{2,1,3}$	$^1/_2\sin\theta_{2,1,4}$	$^1/_2\sin\theta_{2,1,5}$	$^1/_2\sin\theta_{1,2,6}$	
$^1/_2\cos\theta_{3,1,4}$	$^1/_2\cos\theta_{3,1,5}$	$^1/_2\cos\theta_{4,1,5}$		
$^1/_2\sin\theta_{3,1,4}$	$^1/_2\sin\theta_{3,1,5}$	$^1/_2\sin\theta_{4,1,5}$		
$^1/_4\cos\phi_{3,1,2,4}$	$^1/_4$ co s$\phi_{3,1,2,5}$	$^1/_4\cos\phi_{4,1,2,5}$		
$^1/_4\cos\phi_{6,2,1,3}$	$^1/_4\cos\phi_{6,2,1,4}$	$^1/_4\cos\phi_{6,2,1,5}$		
$^1/_4\cos(3\phi_{6,2,1,3})$	$^1/_4\cos(3\phi_{6,2,1,4})$	$^1/_4\cos(3\phi_{6,2,1,5})$		

Coordinate System II. Inverse distances are used as weight function coordinates, but the Taylor series expansions involve all possible valence coordinates. All possible valence coordinates can be easily automatically generated from a Cartesian geometry; a valence bond is included for any pair of atoms that were closer than 1.4 times the sum of their covalent radii, a valence bend is included for any pair of bonds that share a common atom, and a dihedral included for any pair of valence bends that share a valence bond. For methanol, the total valence coordinate set is composed of 5 bonds, 7 bends, and 15 dihedrals. We have taken the inverse of the bond lengths, the cosine of the bending angle, and the cosine of the dihedral angles.

Using all valence coordinates only moderately increases the level of chemical intuition but restricts the applicability to bound state problems. Using the inverse distances and the cosine of the angles imposes further assumptions on the likely form of the potential energy surface. This approach is similar to the hybrid scheme used by Rhee et al.[26] in which the inverses were used for the weight function coordinates and functions of the valence coordinates used for the Taylor series expansion. The major difference being that, rather than choosing ($3n_{\mathrm{atom}} -$ 6) coordinates at the outset, we have allowed all possible valence coordinates.

Coordinate System III. The third coordinate system uses weight function coordinates that are based on a **Z**-matrix description but are scaled to more closely reflect their relative ranges. The weight function uses a total of 15 coordinates, made up from 5 bond functions, 7 bending functions, and 3 dihedral functions. Compared to coordinate system II, only some of the valence coordinates are used, and from this smaller set of coordinates more functional forms have been taken (see Table 3). The expansion coordinates consist of 10 bond functions, 14 bending functions, and 9 dihedral functions.

This coordinate system is close in spirit to that used by Nguyen et al.,[21] as we have carefully chosen our weight function coordinates based on what would be used in a **Z**-matrix. Again, the major difference is that many more possible expansion coordinates have been included. One can imagine that by more carefully choosing the expansion coordinates and tailoring the functional forms even better systems coordinates systems could be devised.

Using coordinate system II, an interpolated PES was grown using the sampling and selection scheme described above. The iterative refinement was stopped after 240 data points had been added to the interpolated potential surface. On this interpolated potential surface a metropolis sample of 1000 geometries were generated and the ab initio potential energy calculated. Assuming that the interpolated potential is sufficiently converged, this sample can be taken to be an independent test set that can be used to gauge the errors in the interpolated potential surfaces.

Rather than regrowing the interpolated surfaces for the other coordinate systems the ab initio data associated with the potential surface constructed using coordinate system II (energies, derivatives and Hessians) was reused. Recycling, rather than regrowing, eliminates the inherent variation between the interpolated potential surfaces that arises from the random sampling, making it easier to compare the convergence of the different potential surfaces.

To demonstrate what effect the coordinate system has on the accuracy of the interpolated PES, the mean absolute error over the test set was calculated as a function of the number of data points. Figure 1 shows the convergence of the interpolation error with the number of data points for coordinate systems I−III. The interpolation error systematically decreases from I to II to III, which demonstrates that as more chemical intuition is used in choosing of coordinates the accuracy of the interpolated potential surfaces increases. Relative to coordinate system I, coordinate systems II and III reduce the interpolation errors by around 65% and 70%, respectively.

Optimizing the Local Coordinates. In coordinate systems II and III there are many more expansion coordinates than necessary to provide a local description of the potential energy surface; coordinates systems II and III contain 28 and 33 expansion coordinates, respectively, whereas a minimum of 12 coordinates are needed to describe methanol. Figure 2 shows an example of the optimization for a data point using coordinate system III. The error (relative to that obtained using all 33 expansion coordinates) is plotted as a function of the number of changes made to the coordinate system. The changes to the coordinate system are either the addition or the removal of a single coordinate and correspond to a single iteration in the optimization. Three initial guesses were used and the initial and final number of coordinates is shown for each optimization run. In Figure 2 it can be seen that the error associated with a given Taylor series can be significantly reduced by optimization of the expansion coordinates.

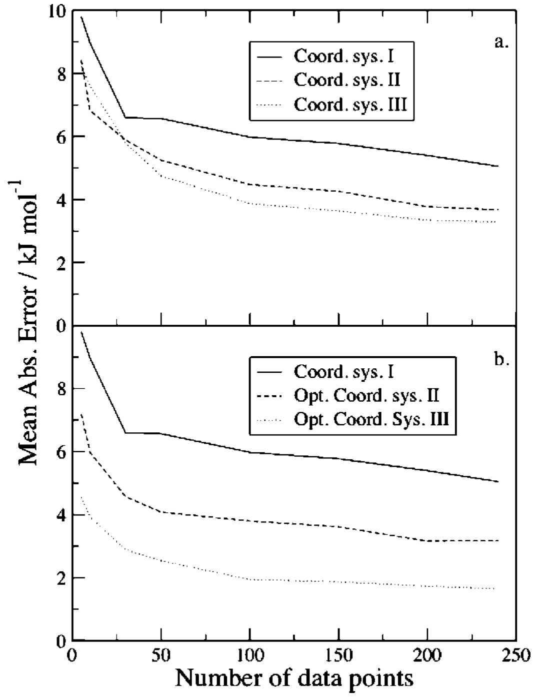

Figure 1. Convergence of the interpolation errors for coordinates systems I−III as a function of the number of data points. Upper panel: all coordinates used in each local expansion. Lower panel: for coordinate systems II and III, the coordinates used in each of the expansions optimized by minimizing the interpolation error over the candidate geometries.

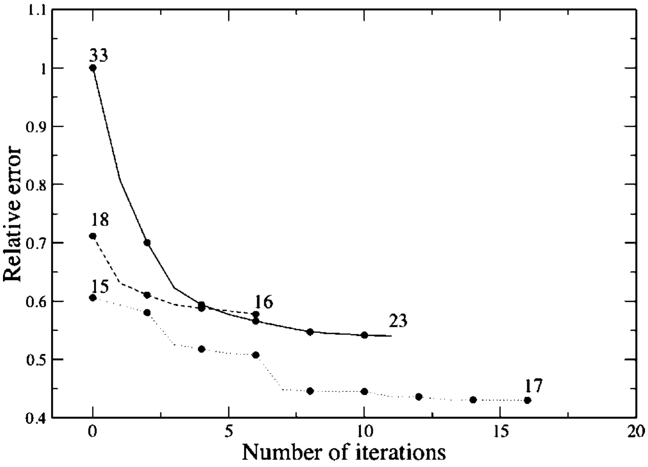

Figure 2. Eample of the optimization of the coordinates used in a Taylor expansion. Coordinate system III was used, and starting from three initial coordintes sets, the interpolation error to neighboring candidate geometries is minimized by changing the interpolation coordinates. The interpolation error is given relative to the error using all 33 coordinates and is plotted as a function of the number of changes to the coordinate set (either adding or removing a coordinate): the solid line started with all coordinates; the dashed line started with valence stretches and cosine of the bending and dihedral angles. The dotted line started with the inverse distances. The curves are labeled with the initial and final number of coordinates.

After optimization, the average number of expansion coordinates is reduced from 28 to 18 for coordinate system II and from 33 to 20 for coordinate system III. This corresponds to about a 50% reduction in the computational cost of a Taylor series evaluation. The convergence of the interpolation errors for the optimized coordinate systems II and III is shown in Figure 1b. After optimization, the accuracy of the interpolated potential surfaces is significantly improved and the difference between coordinates systems II and III is increased. Optimization of coordinate systems II and III reduces the interpolation errors to 60% and 30% of those from coordinate system I.

The larger increase in accuracy of coordinate system III over coordinate system II upon optimization can be understood in

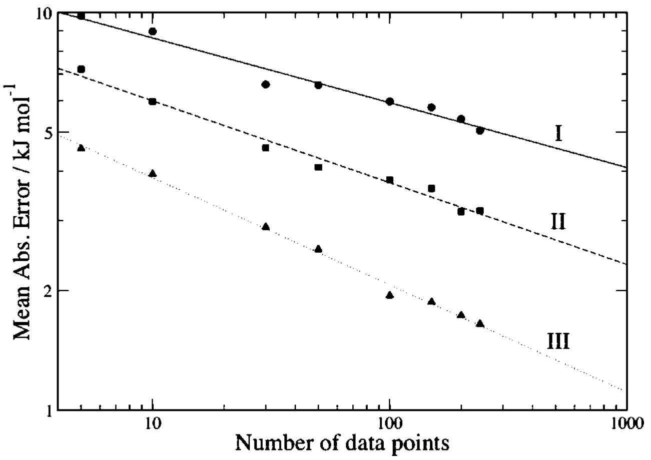

Figure 3. Comparison of the rates of convergence of the interpolation error, with respect to the number of data points, for coordinate system I and optimized coordinate systems II and III. The data are fitted to a power law, $An_{\mathrm{data}}^{-3/f}$, the parameters of which are given in Table 3.

TABLE 4: Parameters of the Fit to the Data from Figure 3[a]

potential	A/kJ mol^{-1}	f
I	12.6	18.4
II-opt	9.6	14.6
III-opt	7.2	11.1

[a] The data were fitted to $An_{\mathrm{data}}^{-3/f}$, where A is the average error in single Taylor series and f the effective dimension.

terms of the number and type of coordinates in each system. Although coordinate system III contains better suited expansion coordinates, as there are more coordinates in coordinate system III than in coordinate system II, this advantage cannot be fully realized unless the number of expansion coordinates in the Taylor series is reduced.

To better understand the differences in the convergence of the interpolated PESs using different coordinates systems, Figure 3 presents the interpolation error versus the number of data points using a log−log scale. The error in a single Taylor expansion increases as the third power of the distance from the data point, and the distance between data points decreases as the number of data points increases. The distance between data points, and hence the distance to the nearest data point, depends on the volume of the system. Taking these factors into account, the average interpolation error varies as $An_{\mathrm{data}}^{-3/f}$. Here A corresponds to the interpolation error for an interpolated PES with a single data point, and f the effective dimension of the configuration space.

The coefficients were determined by linear regression of the log−log data and are listed in Table 4. The decreasing size of A on going from coordinate systems I to II to III indicates that the accuracy of a single Taylor series is improved, which results in an increase of the overall interpolation accuracy.

We also observe that the effective dimension appears to decrease on going from coordinate system I to III. The effective dimension that is obtained from the log−log data can be different from the actual dimension ($3n_{\mathrm{atom}} - 6$). The coordinates that were used in the weight function or in the Taylor series expansions can influence the effective dimension. For example, weight function or expansion coordinates that are able to exploit any approximate separability in the PES will lower the apparent dimension. Conversely, the weight function coordinates may misjudge the distance between molecular geometries, for example, by over emphasizing the role of dihedral coordinates. A fundamental assumption of the interpolation scheme is that there is a correlation between distance to a data point and the accuracy of the Taylor series; if this does not hold then many more data points are required to “fill in” the mistakes which leads to an increase in the effective dimension.

Coordinate systems I and II both use inverse distances as weight function coordinates and both have effective dimensions that are greater than the actual dimension of methanol (18.4 and 14.6 vs 12). Furthermore, coordinate system III, which has carefully tuned weight function coordinates, has an effective dimension that is slightly less than the actual of dimension of methanol (11.1 vs 12). These results suggest that in some situations the inverse distances are ineffective at measuring distance, and by carefully selecting weight function coordinates this problem can be avoided. In addition, optimization of the expansion coordinates improves the accuracy of the interpolated potential surface.

Using coordinate system III, for 240 data points the mean absolute interpolation error is around 1.6 kJ mol^{-1} for a range of energies of 260 kJ mol^{-1}. At the outset, coordinate system III is about twice as accurate as the traditional inverse distance interpolation and, due to the difference in apparent dimension, by 240 data points coordinate system III is around 5 times more accurate. By extrapolating from these results, we see that these modifications to the interpolation scheme result in an appreciable improvement in accuracy; to achieve chemical accuracy (1 kJ mol^{-1}), coordinate system III is estimated to require 1500 data points and coordinate system I might need in excess of 5 million data points. Note that an interpolation error of 1.6 kJ mol^{-1} (134 cm^{-1}) appears to be too large for useful calculation of spectroscopic quantities such as vibration−rotation transition energies. As Figure 3 indicates, this error can be reduced by the addition of more data points. However, an investigation of the convergence of the calculated vibrational levels with the number of data points would need to be performed.

Concluding Remarks. We have detailed modifications to the modified Shepard interpolation scheme that improve the accuracy of interpolated PESs. The type and number of coordinates used to express the potential energy have been increased to include functions of valence coordinates, although in principle any desired functional form could be employed. The scheme presented then uses a training set (a set of geometries at which the ab initio energy is known) to choose the best set of local coordinates at each data point. Three coordinate systems that correspond to increasing levels of flexibility were investigated. Coordinate system I used inverse distances as weight function and expansion coordinates. Coordinate system II retained inverse distances as weight function coordinates but allowed the expansion coordinates to include functions of all possible valence coordinates. Coordinate system III used functional forms for the weight function and expansion coordinates that were chosen from the usual types of **Z**-matrix coordinates. It was found that increasing the amount of chemical intuition systematically reduced the interpolation errors and increased the rate at which the interpolation errors converge with the number of data points. The improvements to the interpolation scheme significantly reduce the interpolation errors and therefore allow larger systems to be treated.

Appendix 1: Modification of the Metropolis Sampling

In Metropolis sampling, starting at a geometry $\mathbf{X}$ with energy E, the molecule is randomly displaced to $\mathbf{X} + \delta\mathbf{X}$, which has energy E'. This step is accepted if a randomly generated number, $0 < \zeta < 1$, satisfies

$$\zeta < \exp[-\beta(E' - E)]$$

The parameter, $\beta = 1/kT$, determines the distribution of the energies in the sample; larger values of β bias the sample to lower energies. The displacement process is repeated for a number of steps until it can be assumed that the geometry has been sufficiently randomized. For an efficient sampling, the size of $\delta\mathbf{X}$ should be chosen so that the percentage of accepted steps is neither too high nor too low.

Problems arise when metropolis sampling is applied to systems that have modes with very different characteristics. Methanol is an example of such a system as it has very stiff bond stretching and bending modes and a very loose torsional mode. From this it is apparent that there is a separation of length scales, as changes in the bond lengths are small in comparison to the changes in the torsional modes. For a random displacement of a fragment to be accepted, the step size must be of the order of an acceptable bond length change; however, this leads to very inefficient sampling of the torsional mode.

Similarly, there is also a separation of energy scales. The stiff modes contain a large amount of zero point energy, much more than the energy in the torsional mode. The metropolis algorithm will evenly spread the total energy between the modes leading to a physically unrealistic sampling of the geometries.

To address these concerns, we have sampled the stiff vibrational modes and the torsional mode with different step sizes and temperatures. The bound state vibrational motion of a molecule can be divided into a number of subsystems, each corresponding to a functional group. The geometry is given by $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{L}\mathbf{X}_N\}$, where $\mathbf{X}_i$ is the geometry of the ith functional group. The internal motions of each functional group are only bond stretching and bond bends and the inverse temperature, β_i, is related to the zero-point energy of these modes. The step size $\delta\mathbf{X}_i$ for each fragment is chosen to ensure that the rate of acceptance is reasonable. There are also a β and $\delta\mathbf{X}$ that describe the energy and length scales of the torsional modes. Rather than taking a random Cartesian displacement, the torsional displacement is a random rotation of the molecule in the corresponding torsional modes. The displacement step in the metropolis sampling is performed as follows:

For each functional group i in the molecule:

1. At the current geometry $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{L}\mathbf{X}_N\}$, the total energy E is calculated.

2. A random Cartesian displacement of the ith functional group is performed, $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{L}, \mathbf{X}_i + \delta\mathbf{X}_i, \mathbf{L}, X_N\}$, and the total energy E' calculated.

3. The step is accepted if $\xi < \exp[-\beta_i(E' - E)]$.

For each torsional mode, i, of the molecule:

1. At the current geometry $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{L}\ \mathbf{X}_N\}$, the total energy E is calculated.

2. A rigid rotation of the ith torsional mode of size $\delta\mathbf{X}$ is performed, and the total energy E' calculated.

4. The step is accepted if $\xi < \exp[-\beta(E' - E)]$.

Appendix II: Optimization of Local Coordinates

Here it is outlined how from a large set of coordinates $\mathbf{Z}^s$, a smaller set of expansion coordinates, $\mathbf{Z}^n$, is selected. The aim is to minimize the error in the expansion for the potential energy over a set of nearby geometries at which the ab initio potential energy is known. The set of geometries near the nth data point is denoted $\{\mathbf{X}(i)\}_{i=1}^{n_\text{samp}}$ and the ab initio potential energies $\{E(i)\}_{i=1}^{n_\text{samp}}$. As the error in the Taylor expansion from the nth data point to the sample point $\mathbf{X}(i)$ is proportional to the cube of the distance, the errors in the expansion are divided by the cube of the distance. As the energy distribution of the sample geometries may not match the desired energy distribution, the errors are also weighted by the probability of the energy occurring in the desired energy distribution

$$\text{err}(\mathbf{Z}^n) = \sum_{i=1}^{n_\text{samp}} \text{prob}[E(i)] \frac{|T(\mathbf{Z}^n;n) - E(i)|}{\|\mathbf{Z}^w[\mathbf{X}(n)] - \mathbf{Z}^w[\mathbf{X}(i)]\|^3}$$

This error is closely related to the confidence radius that was previously introduced with the two-part weight function.[35] The confidence radius is the distance beyond which the rms error exceeds some value. By choosing $\mathbf{Z}^n$ so that err($\mathbf{Z}^n$) is minimal, we are in effect maximizing the confidence volume of the nth data point.

As the problem of choosing the local coordinates is combinatorial in nature, there are a very large number of possible subsets that could be chosen. As there are too many possible coordinate systems to perform an exhaustive exploration, we have chosen to use a discreet optimization approach. First it is necessary to define what are neighboring coordinate sets in $\mathbf{Z}^s$. The coordinate sets $\mathbf{Z}' \subset Z^s$ and $\mathbf{Z}'' \subset Z^s$ are neighbors if they differ by a single coordinate. The optimization algorithm is as follows:

0. Given a starting set of coordinates, $\mathbf{Z}' \subset \mathbf{Z}^s$, the error is calculated and stored as E^min.

1. If none of the neighbors of $\mathbf{Z}'$ have an error smaller than E^min then optimization is complete and $\mathbf{Z}'$ is returned.

2. Otherwise, the neighbor of $\mathbf{Z}'$ with the smallest error is set to be $\mathbf{Z}'$, E^min is updated and we return to step 1.

The above algorithm is a discrete version of steepest descent; from a starting point, head downhill, and stop when a configuration is reached for which every direction is uphill.

As with any steepest descent approach, it is possible to converge to a local minimum. As it is well-known that for a large molecule there is no optimal choice for the **Z**-matrix, we expect there to be many coordinate choices that have a similar error. This problem is addressed by taking a number of starting guesses and taking the one that results in the smallest error. Although this is not guaranteed to result in the optimal coordinate choice, this ambiguity does not present a problem to the interpolation scheme.

References and Notes

(1) Murrell, J. N.; Carter, S.; Frantos, S.; Huxley, P.; Varandas, A. J. C. *Molecular Potential Energy Functions*; John Wiley & Sons Inc.: New York, 1984; p 206.

(2) Laganà, A.; Riganelli, A. *Reaction and Molecular Dynamics*; Proceedings of the European School on Computational Chemistry, Perugia, Italy, July 1999; Springer, Berlin, 2000; p 312.

(3) Jordan, M. J. T.; Gilbert, R. G. *J. Chem. Phys.* **1995**, *102*, 5669–5682.

(4) Kuhn, B.; Rizzo, T. R.; Luckhaus, D.; Quack, M.; Suhm, M. A. *J. Chem. Phys.* **1999**, *111*, 2565–2587.

(5) Sharma, A. R.; Wu, J.; Braams, B. J.; Carter, S.; Schneider, R.; Shepler, B.; Bowman, J. M. *J. Chem. Phys.* **2006**, *125*, 224306.

(6) Huang, X.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2006**, *110*, 445–451.

(7) Brown, A.; Braams, B. J.; Christoffel, K.; Jin, Z.; Bowman, J. M. *J. Chem. Phys.* **2003**, *119*, 8790–8793.

(8) Ho, T.-S.; Rabitz, H. *J. Chem. Phys.* **1996**, *104*, 2584–2597.

(9) Ramachandran, B.; Peterson, K. A. *J. Chem. Phys.* **2003**, *119*, 9590–9600.

(10) van der Avoird, A.; Pedersen, T. B.; Dhont, G. S. F.; Fernandez, B.; Koch, H. *J. Chem. Phys.* **2006**, *124*, 204315.

(11) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. *J. Chem. Phys.* **1995**, *103*, 4129–4137.

(12) Brown, D. F. R.; Gibbs, M. N.; Clary, D. C. *J. Chem. Phys.* **1996**, *105*, 7597–7604.

(13) Lorenz, S.; Gross, A.; Scheffler, M. *Chem. Phys. Lett.* **2004**, *395*, 210–215.

(14) Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. *J. Phys. Chem. A* **2006**, *110*, 5295–5304.

(15) Malshe, M.; Raff, L. M.; Rockley, M. G.; Hagan, M.; Agrawal, Paras M.; Komanduri, R. *J. Chem. Phys.* **2007**, *127*, 134105.

(16) Maisuradze, G. G.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2003**, *119*, 10002–10014.

(17) Maisuradze, G. G.; Kawano, A.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2004**, *121*, 10329–10338.

(18) Guo, Y.; Harding, L. B.; Wagner, A. F.; Minkoff, M.; Thompson, D. L. *J. Chem. Phys.* **2007**, *126*, 104105.

(19) Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080–8088.

(20) Jordan, M. J. T.; Thompson, K. C.; Collins, M. A. *J. Chem. Phys.* **1995**, *102*, 5647–5657.

(21) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522–5530.

(22) Ishida, T.; Schatz, G. C. *J. Chem. Phys.* **1997**, *107*, 3558–3568.

(23) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.

(24) Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 024104-10.

(25) Evenhuis, C. R.; Manthe, U. *J. Chem. Phys.* **2008**, *129*, 024104.

(26) Rhee, Y. M.; Lee, T. G.; Park, S. C.; Kim, M. S. *J. Chem. Phys.* **1997**, *106*, 1003–1012.

(27) Dzegilenko, F.; Qi, J.; Bowman, J. M. *Int. J. Quantum Chem.* **1998**, *65*, 965–973.

(28) Kim, Y.; Corchado, J. C.; Villa, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718–2735.

(29) Yagi, K.; Taketsugu, T.; Hirao, K. *J. Chem. Phys.* **2002**, *116*, 3963–3966.

(30) Bettens, R. P. A. *J. Am. Chem. Soc.* **2003**, *125*, 584–587.

(31) Wu, T.; Manthe, U. *J. Chem. Phys.* **2003**, *119*, 14–23.

(32) Crespos, C.; Collins, M. A.; Pijper, E.; Kroes, G. J. *J. Chem. Phys.* **2004**, *120*, 2392–2404.

(33) Thompson, K. C.; Collins, M. A. *Faraday Trans.* **1997**, *93*, 871–878.

(34) Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 8302–8316.

(35) Bettens, R. P. A.; Collins, M. A. *J. Chem. Phys.* **1999**, *111*, 816–826.

(36) Press, W. H.; Teukolsky, S. A.; Vetterling, W. A.; Flannery, B. A. *Numerical Recipes in Fortran77: The Art of Scientific Computing*; Cambridge University Press: Cambridge, MA, 1992.

(37) Collins, M. A.; Parsons, D. F. *J. Chem. Phys.* **1993**, *99*, 6756–6772.

(38) Godsi, O.; Evenhuis, C. R.; Collins, M. A. *J. Chem. Phys.* **2006**, *125*, 104105.

(39) Evenhuis, C. R.; Manthe, U. *J. Chem. Phys.* **2008**, *129*, 024104.

(40) Moyano, G. E.; Collins, M. A. *J. Chem. Phys.* **2004**, *121*, 9769–9775.

(41) Pulay, P.; Fogarasi, G. *J. Chem. Phys.* **1992**, *96*, 2856–2860.

(42) Baker, J.; Kinghorn, D.; Pulay, P. *J. Chem. Phys.* **1999**, *110*, 4986–4991.

(43) Baker, J.; Pulay, P. *J. Comput. Chem.* **2000**, *21*, 69–76.

JP8103722