

**Interpolating Moving Least Squares Methods for Fitting Potential Energy Surfaces:
Detailed Analysis of One-Dimensional Applications**

Gia G. Maisuradze and Donald L. Thompson

Department of Chemistry, Oklahoma State University, Stillwater, OK 74078

Albert F. Wagner

Chemistry Division, Argonne National Laboratory, Argonne, IL 60439

Michael Minkoff

*Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL
60439*

Abstract

We present the basic formal and numerical aspects of higher-degree interpolated moving least squares (IMLS) methods. For simplicity, applications of these methods are restricted to two 1-D test cases: a Morse oscillator and a 1-D slice of the $HN_2 \rightarrow H + N_2$ potential energy surface. For these two test cases, we systematically examine the effect of parameters in the weight function (intrinsic to IMLS methods), the degree of the IMLS fit, and the number and placement of potential energy points. From this systematic study, we discover compact and accurate representations of potentials and their derivatives for first-degree and higher-degree (up to nine-degree) IMLS fits. We show how the number of ab initio points needed to achieve a given accuracy declines with the degree of the IMLS. We outline automatic procedures for ab initio point selection that can optimize this decline.

I. Introduction

The development of accurate potential energy surfaces (PESs) that are determined from *ab initio* calculations is still a major issue in theoretical studies of chemical reaction dynamics. In spite of the prospects for straightforward *ab initio* dynamics simulations,¹ there is still a need to develop better methods for fitting analytic PESs. The levels of electronic structure theory required to make *ab initio* dynamics feasible are often inadequate for reactions. Even with relatively inexpensive electronic structure methods, multiple studies on a PES (e.g., dynamics, kinetics, mechanisms, and isotope effects) make global fits to all or to large portions of the PES in principle advantageous. However, globally fitting *ab initio* PESs is still more an art than a science. Even if the fitting procedure were routine, there is also a scaling problem with respect to PES dimensions. For N atoms, there are $3N-6$ internal degrees of freedom. If m *ab initio* points are needed on the average to independently fit each degree of freedom, then the fully coupled potential requires m^{3N-6} *ab initio* points to establish the global fit. This high dependence on N means that in practice most global fits to *ab initio* PESs are for triatomic systems.^{2,3} Relatively straightforward, “artless” fitting methods with weak scaling properties with respect to PES dimensions would find ready use in many reaction dynamics simulations.

During the past 30 years, and especially in the past decade, a variety of PES fitting methods have been developed.³ These methods can be categorized as weighted or unweighted. Unweighted methods include least squares fitting methods,⁴ spline methods,⁵ reproducing Kernel Hilbert space (RKHS) interpolation methods,⁶ and hybrid methods

such as Morse-spline⁷ and rotated Morse-spline.⁸ With such methods, the fit at each PES geometry is determined by all the ab initio points. Weighted methods include the modified Shepard interpolation method^{9,10,11,12} and the interpolated moving least squares (IMLS) method.^{13,14,15} With such methods, weights assigned to ab initio points make the fit at each PES geometry sensitive to only local ab initio data.

Among the methods mentioned above, the IMLS is the least-used fitting method, though these methods are well known. Three decades ago McLain¹³ explored two-dimensional fits for some simple nonchemical functions using zero-, first-, second-, third- and fourth-degree IMLS methods. The popular modified Shepard interpolation⁹ introduced by Ischtwan and Collins a decade ago is based on the Shepard method itself, which is a zero-degree IMLS method (ZD-IMLS). Recently, Ishida and Schatz¹⁴ presented a variation on the modified Shepard method that incorporates an indirect second-degree IMLS (SD-IMLS) fit. The reason the Shepard method must be modified for chemical applications is that any ZD-IMLS method suffers from the flat-spot phenomenon whereby derivatives of the interpolated surface are zero at every data point. The modifications to Shepard interpolation solve this problem by using a Taylor expansion that included derivatives at each data point.¹⁴ The derivative information can be either directly computed as in the standard modification⁹ or approximated by SD-IMLS fits as in Ishida and Schatz.

While modifications of ZD-IMLS methods are powerful, in fact the flat-spot phenomenon is a feature only of zero-degree IMLS. All higher-degree IMLS methods have well behaved first derivatives. Thus, while SD-IMLS methods can be used to approximate derivatives that are then used in a modified Shepard method, first-, second-,

and all higher-degree IMLS methods can in principle be used directly to fit ab initio points, resulting in smooth well-behaved PESs for chemically interesting systems.

In what amounts to an extension of McLain's work to chemically interesting fitting problems, we recently carried out the first higher-degree IMLS fit of a PES.¹⁶ This work is a straightforward first-degree IMLS (FD-IMLS) application to the triatomic PES⁷ for $HN_2 \rightarrow H + N_2$. The resulting fit was smooth, well behaved, and accurate. The results of this effort motivate us to systematically study the properties of higher-degree IMLS fits for chemical applications. This paper is the first installment of that study and is exclusively dedicated to the simplest of fits, that of a one-dimensional potential curve rather than a potential surface. Other papers in our study will focus on the challenge, articulated earlier, of developing higher-degree IMLS methods with weak dependencies on the number of dimensions of PESs. However, this challenge can best be met by understanding at more elemental levels the performance of IMLS methods as a function of degree, parameters, and ab initio data sparsity in the representation of values and derivatives for one-dimensional curves. With this grounding, the expansion of the method to large numbers of dimensions is not difficult to do.

In this one-dimensional study, we will not directly compute ab initio points at discrete locations and attempt to fit them. Rather, we will use analytic functions, sample these functions at discrete locations as if we were doing ab initio electronic structure calculations, fit the discrete set of energies with IMLS methods, and then evaluate the accuracy of the fit everywhere by using our global knowledge of the analytic function. This approach allows a more quantitative analysis of IMLS methods provided the analytic functions selected are representative of the results of ab initio calculations. For

the two analytic functions we use in this study, that is in fact the case. The first function is a modification of a 1-D slice of the PES for $HN_2 \rightarrow H + N_2$ dissociation reaction used in our earlier study.¹⁶ A collinear slice is selected in which the N-N distance r is fixed at $2.0 a_0$ and R , the distance between H and the center of mass of N_2 , is varied. This particular 1-D slice illustrates a reaction path with a barrier and is similar to, but not exactly the same as, the minimum energy path in the dissociation of HN_2 . (To remove minor artifacts from the 1-D slice, we have slightly modified the analytic PES of Koizumi et al.⁷ which is a fit to high-order ab initio electronic structure calculations.¹⁷) The second function is the Morse function that is routinely used to represent uncoupled and dissociative motion in the bonds of diatomic and polyatomic molecules. The selection of values for the three Morse parameters is not critical to our study, but our test case used representative values of 100 kcal/mole for the dissociation energy, $2.0 a_0^{-1}$ for the β parameter, and $2.0 a_0$ for the equilibrium distance.

An outline of this paper is as follows. The IMLS methods are briefly reviewed in Section II. The application of the methods to the two exact functions is presented in Section III. The discussion of results is in Section IV. A summary and conclusions are in Section V.

II. Method

The basic aspects of the IMLS method have been given in a previous paper¹⁶ and in mathematical reference books.¹⁸ The method therefore will be only briefly outlined here. Consider $N + 1$ data points with $\{x_i; i = 0, 1, \dots, N\}$ abscissas and $\{f_i; i = 0, 1, \dots, N\}$

ordinates. Given a polynomial $p(x) = \sum_{i=0}^m a_i x^i$ where $m \leq N$, one can define a weighted least squares expression for the fitting coefficients a_i as a minimization of

$$E_x(p) = \sum_{i=0}^N w_i(x) [p(x_i) - f_i]^2, \quad (1)$$

where $w_i(x)$ are positive weight functions. Unlike the usual weight functions that depend only on x_i , for IMLS methods the weight functions depend also on x , the location on the PES where a fit is required. The logic of the fit demands that the weight functions have relatively large values for x_i close to x and relatively small values for the more distant x_i . For this work, the weight functions will take the form of

$$w_i(x) = \frac{\exp[-(x - x_i)^2]}{(x - x_i)^n + \varepsilon}. \quad (2)$$

This form has two parameters, ε and n , whose optimal selection will be considered in the next section. While there are a variety of ways of formulating weights,^{10,18} those as expressed in Eq. (2) are representative of IMLS applications.

The usual minimization procedure produces normal equations that are conveniently expressed in matrix-vector form:

$$\mathbf{B}^T \mathbf{W}(x) \mathbf{B} \mathbf{a} = \mathbf{B}^T \mathbf{W}(x) \mathbf{f}, \quad (3)$$

where \mathbf{a} and \mathbf{f} are column vectors, B is an $N \times (m+1)$ matrix, B^T is a transpose matrix, and $W(x)$ is an $N \times N$ diagonal matrix. They have the following form:

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_N \end{bmatrix}, \quad B = \begin{bmatrix} 1 & x_0 & \cdots & x_0^m \\ 1 & x_1 & \cdots & x_1^m \\ & & \ddots & \\ 1 & x_N & \cdots & x_N^m \end{bmatrix} \quad (4)$$

$$W(x) = \text{diag}[w_0(x), w_1(x), \dots, w_N(x)].$$

The solution \mathbf{a} to Eq. (3) provides the coefficients to the fitting polynomial p at point x .

For the $m = 0$ case, the solution to Eq. (3) reduces to

$$a_0(x) = \frac{\sum_{i=0}^N w_i(x) f_i}{\sum_{i=0}^N w_i(x)}. \quad (5)$$

This zero-degree IMLS (ZD-IMLS) solution is known as the Shepard method. When

$m = 1$, the first-degree IMLS (FD-IMLS) solution to Eq. (3) reduces to

$$a_0(x) = \frac{\sum w_i(x) f_i}{\sum w_i(x)} - \frac{\sum w_i(x) x_i}{\sum w_i(x)} a_1 \quad (6)$$

$$a_1(x) = \frac{\sum w_i(x)x_i f_i - \frac{[\sum w_i(x)x_i][\sum w_i(x)f_i]}{\sum w_i(x)}}{\sum w_i(x)x_i^2 - \frac{[\sum w_i(x)x_i]^2}{\sum w_i(x)}} .$$

Second-degree (SD), third-degree (TD), and higher-degree IMLS do not have simple analytic solutions, and numerical methods will have to be used. The most straightforward way involves a matrix inverse. However, since the matrix B, known as a Vandermonde matrix, is ill-conditioned,¹⁸ this approach can be unstable. Singular value decomposition (SVD) is a computationally more expensive numerical approach, but one with dramatically improved numerical stability. We discuss the performance implications of each method in later sections of the paper.

For classical trajectory applications, the derivative of the PES is required. Since IMLS methods use weights that are dependent on location, the derivative of the fit is not a simple matter of taking the derivative of the polynomial fitting function for a constant set of polynomial coefficients. However, the numerical methods to determine the polynomial coefficients **a** can be reused with great efficiency to determine the appropriate derivatives of **a**.^{14,15,18} To illustrate, one can re-express Eq. (3) in the form

$$\mathbf{C} \mathbf{a} = \mathbf{d} , \tag{7}$$

where **a** and **d** are vectors; **C** is a matrix; and **a**, **C**, and **d** all depend on *x*. Taking the derivative of the equation produces the following result:

$$\mathbf{C} \mathbf{a}' = \mathbf{d}' - \mathbf{C}' \mathbf{a} . \quad (8)$$

This is Eq. (7) with a different right-hand-side vector. All the matrix operations on \mathbf{C} by either the matrix inverse or SVD methods can be reused with this different vector. The derivatives used in this paper are done that way. No finite differencing is used.

III. Results

As mentioned in the Introduction, the higher-degree IMLS methods will be evaluated with two test cases: the Morse function and a 1-D slice of PES for the $HN_2 \rightarrow H + N_2$ dissociation reaction developed by Koizumi et al.⁷ Figure 1 gives a plot of these two curves as a function of the distance R for the entire range of R over which the fit will be measured. We will identify these two test cases by the acronyms MO and HN_2 , respectively. As can be seen from Fig. 1, these two test case encompass both dissociative and reaction barrier features that are controlling PES elements in reaction dynamics. Both test cases are sampled with discrete numbers of points in the IMLS fit procedure. However, as mentioned in the Introduction, the evaluation of the fit will be with respect to the entire potential curve displayed in Fig. 1. Both fitting errors displayed as a function of R and as a root mean square (rms) measure will be used in the evaluation. Note that the ~ 100 kcal/mol range for both test cases in Fig. 1 implies that a relatively large 1 kcal/mol rms error still implies a 1% relative error.

As described in the previous section, an IMLS fit is a function of two weight function parameters (ϵ and n), the degree, the number of data points N , and implicitly the

location of those data points. In this section, all these dependencies are examined. While to some extent these dependencies are coupled together, each element is emphasized in turn.

A. Dependence on ϵ

The parameter ϵ controls the deviation from singularity of the weight function when evaluated at an ab initio point. Figure 2 illustrates the rms fitting error as a function of ϵ for the Shepard, FD-IMLS, and SD-IMLS fits for both the MO and NH₂ test cases. For the results in the figure, n is fixed at 6 and N is fixed at 33 for equally spaced ab initio points. Four conclusions can be drawn from the figure. First, the results are qualitatively the same for both test cases. Second, if ϵ is too large, the accuracy of all fits is degraded. Third, the Shepard fit shows a qualitatively different behavior from the fits of FD-IMLS and SD-IMLS in that it has a minimum in the rms error. For FD-IMLS and SD-IMLS fits, as ϵ decrease the rms error reaches a minimum that essentially persists for all further decreased ϵ . Thus, for higher-degree IMLS methods, it is necessary only that ϵ be small enough for optimally accurate fits. As the figure shows, how small is small enough is not a strong function of the degree of IMLS. The fourth conclusion to draw from the figure is that with fixed value of ϵ , increasing the degree of IMLS does decrease the rms fitting error. All the results shown in the figure do not dramatically change with n . Higher degrees of IMLS than displayed in the figure continue the same trends to a lower and lower rms fitting error at smaller values of ϵ .

The results in Fig. 2 do change as a function of N . For example, when $N = 17$, Fig. 3 is analogous to Fig. 2 for both the MO and HN_2 test cases. As can be seen from the two figures, as N increases, the minimum in the Shepard fitting error becomes deeper, narrower, and located at smaller values of ε . For higher-degree IMLS fits, the two figures show that as N increases, the range of optimal values of ε contracts to smaller values.

Figure 4 is analogous to Fig. 2 only for the first derivative, rather than the value, of the potential energy. The scale of the figure is much larger because the range of the first derivative is much larger than the range of energy values for the two test cases. In addition to the scale change, Figs. 2 and 4 have several qualitative differences in the ε dependence they display. The Shepard method for either case has two minima in the rms error for the derivative. One corresponds to the minimum in Fig. 2 for the rms error for the value. However, the second and global minimum in the derivative error occurs at a value of ε where the value error is about a factor of ten above optimum. Hence, accurate values are accompanied by degraded derivatives. This situation implies an oscillatory fit about the true value of the potential whose frequency and amplitude changes with ε . Beyond the Shepard fit, higher-degree IMLS fits show a decreasing sensitivity to ε . Each of the two curves for FD-IMLS has only one minimum, which is near the optimal value of ε for reducing the value rms error. The SD-IMLS has a less pronounced minimum. Higher degree tests show a continuation of this trend of higher-degree IMLS methods producing smoother fits where agreement with derivative and value go hand in hand.

For the HN_2 case and for $N = 33$ and $n = 6$, Fig. 5 displays fit smoothness by plotting the derivative as a function of R for the Shepard method in panel (a) and for FD-

IMLS in panel (b). In each panel, the same three values of ε are used, corresponding to the smallest value in Fig. 4, the value at the minimum for FD-ILMS, and the global minimum for the Shepard method. Also in each panel the exact derivative is displayed. Panel (a) shows that decreasing the value of ε increases the frequency and the amplitude of oscillations in the Shepard derivative. This trend culminates at the lowest values of ε with extreme oscillations that bottom out at a zero value of the derivative in the vicinity of each ab initio point. This is the “flat spot” phenomenon. At some intermediate value of ε , the oscillating fitted derivative is in modest agreement with the oscillating exact derivative. In panel (b), decreasing the value of ε increases the frequency but not the amplitude of oscillations in the FD-IMLS derivative. All small values of ε result in a fitted derivative that closely approximates the exact derivative but might oscillate about it. There is no flat spot phenomenon. Higher-degree IMLS generally result in smoother and more accurate fits to derivatives for small enough ε .

The results in Fig. 5 show that the Shepard method behaves in a qualitatively different way from higher-degree IMLS methods in obtaining derivatives. The origin of this difference is straightforward. For all IMLS fits except the Shepard, the derivative of the fit comes from two contributions: the derivative of the polynomial fit itself and a derivative that reflects how the positional dependence of the weights changes the polynomial fits. Since the Shepard fit is a zero-degree IMLS, for this IMLS fit alone the first contribution to the derivative is zero. Consequently, the entire Shepard derivative is due to positional dependence of the weights. No matter what form the weights take, in practice accurate fits to the value require highly localized weights with correspondingly large and variable derivatives. The result is sensitive Shepard derivatives that are

insufficiently anchored to the ab initio data. As mentioned in the introduction, the solution to this problem in the past has been to use the Shepard method to separately and directly fit values, derivatives, and Hessians. As indicated in Fig. 5b, higher-degree IMLS methods are another way to correct the Shepard limitations for derivatives.

B. Dependence on n

The parameter n sets the width of the weight function about the point at which the potential is desired. Figure 6 is analogous to Fig. 2, only for n varied and ϵ selected at its optimal value. A comparison of Figs. 2 and 6 show that the n dependence of the rms fitting error is much less severe than that on ϵ . All methods behave in qualitatively similar ways with optimal fits occurring for large enough values of n . For any given value of n , the fitting error improves with the degree of the IMLS fit. Although it will not be shown, the rms fitting error for the first derivatives as a functions of n for optimal values of ϵ is similar to Fig. 6. For all fitting methods, large variations in n have little effect on the derivative fitting error once n is large enough.

C. Dependence on N and on Degree

The results in this section so far indicate that optimal fits for different-degree IMLS can be achieved for large enough values of n and for small enough values of ϵ (except for the Shepard method). With optimal values of ϵ and for $n = 6$, the rms fitting error as a function of N for equally spaced points is displayed in Fig. 7 for a selection of

different degrees up to the ninth degree. Figure 7a is for the HN_2 test case, while Figure 7b is for the MO test case. Since N must be larger than the degree, higher-degree IMLS fits are not plotted down to the lowest values of N . As a standard of comparison, Figure 7 also plots the rms fitting error for a cubic spline. Over the full range of N , only discrete values of N are examined in the figure, namely, those that double the grid of equally spaced points starting from a coarsest grid of three points that exactly span the length of R seen in Fig. 1 for each case. Thus $N = 2^m + 1$, where $m = 1, 2, \dots$. In the context of equally spaced points, only this choice of sampling the range of N allows reuse of all previously calculated ab initio points in a process of refining the fit with more ab initio information. The consequences of a strategy of unequally spaced points will be discussed in the next part of this section. The fitting errors in Fig. 7 for equally spaced points are best discussed starting first with the MO case followed by the HN_2 case.

In Fig. 7b, the MO rms error almost always decreases as the degree of IMLS increases. As a consequence, for the larger numbers of data points, many order of magnitude improvements are possible over the Shepard method. At the ninth and highest degree examined, the accuracy degrades with increased degree at the highest number of points. The origin of this change is the beginning of oscillations between data points that high-order power series expansions can suffer from. However, IMLS is not simply a power series because the positional dependence of the weights introduces nonlinear flexibility. This is illustrated by the comparison of IMLS fits to the cubic spline fit displayed in Fig. 7b. Although the zero-degree and first-degree IMLS fits have larger errors than the cubic spline, second- and all higher-degree IMLS fits are always superior to the spline. In effect the nonlinear flexibility of the weights makes the nominal second-

degree power series fit of SD-IMLS more accurate than the third degree power series fit of the spline.

In Fig. 7a for HN_2 , the dependence on degree is less regular than for the MO case. To begin with, no IMLS fit is in general superior to the spline fit. The reason is that the Koizumi et al. HN_2 potential curve we are using is itself a spline fit. The only reason the spline fit in Fig. 7a does not have zero rms error is that the spacing of ab initio points we are using is not identical to the spacing Koizumi et al. used to set the spline. While in principle at any given equally spaced set of N points an IMLS fit could be superior to a spline fit in representing a spline curve based on a different selection of points, in practice the results in Fig. 7b show such occurrences for only at $N = 9$ for the higher-degree IMLS fits. More typically, increasing the IMLS degree up to about the fourth degree decreases the error down to that of the spline, while further increases largely leave the error unchanged except at the highest values of N where the error degrades for the seventh and higher degrees. The origin of this degradation is the oscillations of higher-degree power series expansions discussed in the MO case. The degradation is more severe here because, unlike the MO case, the potential being fit is inherently a low-degree polynomial.

Figure 8 is analogous to Fig. 7 only for the first derivative, rather than the value, of the potential energy. The scale of the rms error is much larger in Fig. 8 because the range of the first derivative is much larger than the range of energy values for the two test cases. Despite the scale change, Figs. 8 and 7 are qualitatively similar to each other with one exception. For the derivatives, the Shepard method shows a much greater insensitivity to N than do any of the other methods. For reasons discussed in Fig. 5(a),

the form of the Shepard method is not well suited to accurate derivatives. The results for both cases in Fig. 8 show that this limitation in form make the method unable to process additional ab initio information on the value into more accurate fits for the derivative.

The rms errors in Figs. 7 and 8 are a global measure of fitting error. However, the error is not evenly distributed over the full range of R sampled by the fit. In Fig. 9, the difference of the exact potential from the Shepard, FD-, SD-, and TD-IMLS fits in Fig. 7 is displayed as a function of R for $N = 17$ for both test cases. For the MO case in Fig. 9b, the majority of the error is located at $R \leq 2.2 a_0$ for all IMLS degrees. Since the equilibrium position for the MO potential curve is $2.0 a_0$, basically the error in the fit at each degree is located in the inner-wall region of the MO potential where the potential is varying most rapidly and is therefore most difficult to fit. The FD-, SD-, and TD-IMLS fits are qualitatively similar with near-exact agreement at the 17 ab initio points and maximal disagreement in about the middle of the interval between adjacent ab initio points. The Shepard error distribution is different with maximal disagreement occurring at the two ab initio points farthest up the inner wall. The Shepard fit disagreement with ab initio points is largely controlled by the value selected for ϵ . As the optimal value of ϵ used in Fig. 9 is reduced to smaller values, the Shepard fit will progressively come into agreement with the ab initio points but at the cost of greater error in the interval between the points. For the HN_2 case in Fig. 9a, many of these trends repeat, with the addition of a build-up of error by all fits in the barrier region of the potential (see Fig. 1). As in the case of inner-wall behavior for the MO potential, Shepard alone among the fits fails to get excellent agreement with all the ab initio points on the barrier because the optimal

value of ϵ that reduces the global rms error does not simultaneously reduce this local error.

The results in Fig. 9 are true for all values of N . Since the construction of Figs. 7 and 8 is based on repeated halving of an equally spaced grid, increasing N lowers the error primarily because it places at least some of the increased points in regions where the fit is poorest. This is a successful but inefficient way to decrease the error by supplying more ab initio calculations.

D. Dependence on Data Point Location

Under ideal circumstances, one would want to generate an accurate PES with the fewest ab initio calculations possible. The results above indicate that performing ab initio calculations on a regular grid whose grid spacing is progressively halved is not the most efficient way to converge the error of the fit. Are there strategies whereby calculations at a few seed points can lead to automatic selection of additional points that efficiently converge the accuracy of the fit? In a strictly mathematical sense, Rice¹⁹ gives a brief outline of an algorithm for spline least-squares fitting with variable points (knots) that includes automatic point selection. This spline problem is mathematically similar to the IMLS fitting problem we are addressing. He proves that there is no automatic selection procedure for progressively selecting a finite number of additional points that will in any optimal sense converge on the most compact spline representation of the data to a given accuracy. Thus, the development of an optimal IMLS algorithm of automatic point selection is not likely rigorously possible. Rice does suggest that functions that specific

application properties may be treated with a practical algorithm. In fact, numerous studies regarding PES automatic point selection indicate that useful approaches can be developed that are broadly useful if not rigorously successful for all possible problems. These studies include automatic point selection on the basis of maximum variance in the least squares fits,²⁰ on the basis of a statistical confidence radius about each ab initio point,¹⁰ on the basis of neural networks,²¹ and on the basis of maximal differences between contending fits.²² Because the emphasis of this paper is on IMLS fits of different degrees, automatic point selection based on differences between contending IMLS fits of different degrees will be examined here. Other principles of automatic point selection just mentioned could perhaps also be applied with profit to IMLS fits of higher degree.

Given m ab initio seed calculations, IMLS fits of degree $m-1$ can be performed. With the proper choice of weight function parameters discussed above, these fits will be negligibly different for geometries very close to the m seed points (except for Shepard fits which will not be included in this discussion). They will, however, be different in intermediate regions far away from the m seed points. The selection of the next k ab initio point could thus be where they are maximally different. The same procedure could be repeated with now $m+k$ ab initio points. Continued repetition could be terminated when the rms difference between the two fits drops below some input amount. This procedure is defined by the number m of initial seed ab initio points, the number k of additional points added per repetition, and the degrees of the two contending IMLS fits.

Table I presents the results of this automatic point selection procedure for $m = 5$ equally spaced ab initio seed points, for $k = 1$ additional point added per repetition, and

for FD-ZD, SD-FD, TD-SD, and FRD-TD as the contending IMLS fits. Both the HN_2 and MO cases are represented. The results are expressed in terms of the total number of ab initio points needed to have an rms error relative to the exact potential of 4.00, 1.00, or 0.25 kcal/mol for the higher degree of the IMLS pair of fits (e.g., for TD-SD, the rms error is for the TD fit). These results are compared to the FD-exact, SD-exact, TD-exact and FRD-exact combinations, in other words, where the next point selected occurs where the IMLS fit is maximally different from the exact potential. This combination represents the ideal, but unattainable, implementation of this automatic points selection strategy. Finally, these results are also compared to those of Fig. 7 for a grid of equally spaced points whose grid spacing is progressively halved until the desired rms error is reached. The results in Table I show that for HN_2 automatic point selection by FD-ZD contending fits is decidedly worse than equally spaced data points in entire region of rms error, whereas for other contending pairs of IMLS fits it is typically not decidedly superior or inferior to equally spaced data points for rms errors of 1.00 kcal/mol or higher. However, for the highest accuracy rms error of 0.25 kcal/mol, automatic point selection can be decidedly better, avoiding the calculation of on the order of ten ab initio points. For the MO case, such savings occur at the 1.00 kcal/mol rms error level or lower. The difference between contending pairs of IMLS fits and contrasting an IMLS fit with the exact potential for FD- and SD-IMLS is noticeable, especially for rms error of 0.25 kcal/mol, whereas for TD- and FRD-IMLS it is insignificant. Similarly, there are only relatively minor differences between a maximum IMLS degree of three or four. While the results of SD-FD automatic point selection are better than equally spaced points, the number of data points used for the achievement of highest accuracy of rms error is much bigger than

number of data points used in TD-SD and FRD-TD contending fits. Therefore, both FD-ZD and SD-FD are generally not as successful as contending higher-degree IMLS fits. This is especially true for the FD-ZD.

In Table II, variations by one in the value of m and k are displayed for the TD-SD and the FRD-TD pairs in Table I. Increasing the number of equally spaced seed points from five to six usually leaves unchanged or increases the total number of ab initio points needed to get convergence. Larger variations in m tend to show that even values of m typically require more points than odd values of m . Variations in the number of points added at a time never decrease the total number of ab initio points needed and about 25% of the time increase the total number of ab initio points. Not shown in the table is the smallest value of m , that of $m = 3$, which is not a good choice, since for this number of data points TD- and FRD-IMLS methods cannot be applied. In general, although superior to even values of m , five or seven initial seed points display better performance for the two test cases.

The results in Table I do not deal with the termination of the point selection scheme. The table lists the rms error of the fits relative to the exact potential, something that is not knowable in an actual application. This is in fact a general problem for automatic point selection or repeated halving of a uniform grid. For automatic point selection, the rms difference between contending IMLS fits is known, and termination can therefore be based on the reduction of this rms difference below some input value. Under ideal circumstances, one might hope the rms difference between contending fits might approximate to within some simple factor the rms error between the highest-degree IMLS fit of the pair and the exact potential. Then selection of the input value for the rms

difference would be equivalent to selecting the desired final accuracy of the fit. Table III gives the results of this approach for the TD- and FRD-IMLS fits of Table I. As in Table II, the first column gives the desired input value for the rms error of the TD- or FRD-IMLS fit with respect to the exact potential. Three other columns for each IMLS fit are labeled by the ratio of this input value to the input value for the rms difference between contending IMLS fits. This ratio times the desired rms error is used to terminate the automatic point selection. Each entry in the table gives two pieces of information. The first is the total number of ab initio points needed to reduce the rms difference below the input difference value. The second in parentheses gives the rms error of the terminated fit with respect to the exact potential. Thus, an entry in the table for a TD-IMLS column labeled 2.0 (for rms-difference/rms-error) and a row labeled 1.00 (for rms-error) indicates how many ab initio points are required to reduce the rms difference between the TD- and SD-IMLS fits below 2.00 kcal/mol and, in parentheses, the TD-IMLS rms error with respect to the exact potential. The best value of the rms-difference/rms-error ratio is the one that makes the value in parentheses for each entry most like the rms error label for that entry's row.

The results in Table III for HN_2 indicate that setting rms difference input value to half the desired rms accuracy of the fit produces fits with rms errors close to the desired value. The worst mismatch in an absolute sense is a 5.47 kcal/mol rms error when a 4.00 kcal/mol error was desired. The worst mismatch in a relative sense is a 0.38 kcal/mol rms error when a 0.25 kcal/mol error was desired. For the MO case, the results in Table III are more systematic but more complicated. For the highest desired rms error of 4.00 kcal/mol, the rms difference input value should be half that value. For the middle desired

rms error of 1.00 kcal/mol, the rms difference input value should be identical to that value. For the lowest desired rms error of 0.25 kcal/mol, the rms difference input value should be twice that value. With such a selection, the final accuracy of the fit is never more than 0.18 kcal/mol above the desired value. A simpler but less efficient strategy is to set the input rms difference to the value of the rms error desired. Then only for the least accurate fit (4.00 kcal/mol) is the final fit accuracy too large (by at most 1.06 kcal/mol). Importing the best strategy for the HN_2 case to the MO case will result on average in about 20% more ab initio calculations with all fitting errors no more than 5% above the desired rms error value. Overall, the results in Table III suggest that automatic point selection schemes based on contending IMLS fits can be terminated with controllable final accuracy.

The results in Tables I–III apply if one is planning to fit a PES prior to its use. Automatic point selection is carried out under static conditions where an rms error below a specified size is desired everywhere the PES is below some cutoff energy. However, there are other circumstances where automatic point selection is carried out in a more dynamic environment. For example, in ab initio dynamics, a trajectory dictates where information on the PES is needed next. Consider a trajectory that has just dictated the calculation of the Nth ab initio point. The trajectory now proceeds to propagate away from the Nth point on a PES fit constructed from the N ab initio points. If N points were appropriately scattered near the trajectory's intended path and if the PES fit were sufficiently accurate, perhaps the trajectory could go to completion without another ab initio calculation. On the other hand, if the N points were localized away from the intended path of the trajectory, after k steps, the trajectory might eventually wander far

enough away from the N ab initio points so that the PES fit could not be trusted and an $(N+1)$ th ab initio calculation would have to be performed. Following the same ideas discussed in connection with Tables I–III, IMLS fits can be used to automatically determine when a new ab initio calculation must be calculated.

Consider the MO test case under the extreme condition that the trajectory is at the N th point R_N propagating to larger values of R while the other $(N-1)$ points are at smaller values of R . How far beyond R_N , that is, ΔR , can the trajectory proceed before the error in the IMLS fit becomes unacceptable? To be concrete, let R_N be the larger value of R where the MO potential exactly equals 50 kcal/mol (i.e., half its dissociation energy). Let an acceptable error be 0.01 kcal/mol. Let the N previously calculated ab initio points be scattered in a uniform grid between R_N and the inner value of R where the MO potential also exactly equals 50 kcal/mol. If $N=87$, then a FD-IMLS fit gives an rms error between the two R values of less than 0.01 kcal/mol. If $N=19$, then a TD-IMLS fit gives a better than 0.01 kcal/mol rms error. For these conditions, the value of ΔR calculated under three options is displayed in Fig. 10 as a function of the degree of IMLS fit. For each of the two values of N there are four curves. The first is the largest value of ΔR such that the absolute difference between the IMLS fit and the exact MO potential is less than 0.01 kcal/mol. This is an ideal value of ΔR because it is calculated with knowledge of the exact potential, information not attainable in actual applications. The figure shows that this ideal ΔR grows at a little less than exponential in the degree of the IMLS fit (i.e., a nearly straight line in the figure with a downward curvature). The second curve is the largest value of ΔR with a less than 0.01 kcal/mol absolute difference between IMLS fits of the m th and the $(m-1)$ th degree. For a given degree, this value of ΔR is always smaller

than that for the first curve. In other words, the absolute difference between the m th and $(m-1)$ th IMLS fits considerably overestimates the absolute difference between the m th IMLS fit and the exact potential. This overestimation results in a quite conservative estimation of ΔR . The third curve is the largest value of ΔR , with a less than 1.5×0.01 kcal/mol absolute difference in IMLS fits. In the language of Table III, for this curve the ratio rms-difference/rms-error is set to 1.5. This value leads to a closer but uniformly conservative estimate of ΔR . The fourth curve is the largest value of ΔR , with a less than 3.0×0.01 kcal/mol absolute difference in IMLS fits. This is closer to the ideal value of ΔR but can be slightly larger for the lower-degree IMLS fits. In other words, for lower IMLS degrees, estimating ΔR this way can result in errors slightly larger than 0.01 kcal/mol. The general behavior of the four curves is independent of the value of N . However, the value of ΔR increases with N , although less so with IMLS fits of high degree. Large increases in the acceptable error only cause the four curves to shift up together. Selecting ΔR on the basis of the absolute difference between IMLS fits of degree m and $(m-2)$ or between IMLS fits of degree m and $(m-3)$ has negligible effect on the results in Fig. 10.

A careful examination of Fig. 10 suggests an alternative strategy from that just described. For the higher degree IMLS, the ΔR value for the m th degree determined with a rms-difference/rms-error ratio set to 1.0 is similar to the ΔR value for the $(m-1)$ th degree determined from the rms error with respect to the exact curve. In other words, for the $(m-1)$ th-degree IMLS, the m th-degree IMLS is nearly identical to the exact potential. Thus, if the cost of an m th-degree IMLS fit (see the following section) is too expensive relative to an $(m-1)$ th-degree IMLS fit, the higher-degree IMLS fit need be employed

only to determine an accurate ΔR . This strategy could in principle be applied to automatic point selection under the more static circumstances applicable for Table III. However, the results in the table show that this approach would not be an improvement over the strategy already discussed.

Thus, in this part we conclude that, without knowing the exact potential, the strategy of contending IMLS fits allows an automatic selection of where or whether to perform additional ab initio calculations. This applies to static situations where the large portions of the PES are desired before its use or to dynamic situations where the PES is being generated on the fly. The automatic point selection scheme requires some estimation of the ratio of the differences in contending IMLS fits of different degrees to the difference of the highest-degree IMLS with the exact potential. In the two test cases examined, the ratio has relatively simple dependencies and can be selected in conservative ways. While more detailed studies of multidimensional PESs will be necessary to refine this approach and assess its usefulness, these initial results are encouraging.

IV. Discussion

Based on the results of the previous section, this section discusses in turn issues concerning timings, the weights, and the degree of IMLS methods. The section concludes with comments concerning higher-degree IMLS methods relative to other fitting methods.

The time to evaluate an IMLS fit of degree m is the sum of the timings of three distinct steps. First, there is the construction of the matrix and vectors according to Eq.

(7). The timing to do this step goes as $N(m+1)^2$ because each of $(m+1)^2$ elements involves a summation N ab initio points [see Eqs. (3) and (4)]. Second, there is the manipulation of the matrix by inversion or SVD techniques to obtain the polynomial coefficients. The timing for this step goes as $(m+1)^3$, but the prefactor for SVD is twice that of inversion. Although faster, runs on the two test cases under consideration show that inversion can be unstable for extreme values of ε or for higher degrees ($m \geq 4$). Third, the polynomial is evaluated in what amounts to a dot product whose timing goes as m . Actual computer timings performed for an entire IMLS fit evaluation consistently show a time-to-solution dependence that goes as $(m+1)^2$. Detail decompositions of the timings consistently show that the matrix construction step dominates because typically $N \gg m$. This situation is expected to remain true for multidimensional applications. Hence the computational cost of evaluations of higher-degree IMLS fits goes as the square of the degree. Also in this context, the additional cost of SVD over matrix inversion is an insignificant factor in the overall cost of an IMLS evaluation.

Although the basis of an IMLS fit is a polynomial, the location-dependent weights in effect add a nonlinear flexibility. Hence the weights are not a casual but an integral component of the success of the method. Since all weights have parameters, any widespread applicability of IMLS methods will require that near-optimal parameter values can be straightforwardly selected for each application. The results in the previous section show that there is a common range of parameter values that is optimal for both test cases for first-degree and higher IMLS fits. To the degree that the two test cases are representative, this same range should be nearly optimal for a broad range of applications.

Although not examined in the previous section, the weights can also play an important role in the scalability of IMLS methods to PESs with large numbers of internal degrees of freedom. As described in Section II, the matrices that must be assembled for each evaluation of the IMLS fit involve summations over *all* the ab initio points available. This can be a very large number of points for high-dimensional PESs, and this step dominates the time it takes to do an IMLS fit evaluation. Fortunately, the nature of the weights provides opportunities to dramatically reduce this dependence. No matter what the form, weights decline to zero rapidly with the distance of an ab initio point from the point where the IMLS fit is required. For large enough distances, the weights need not be evaluated at all but can be effectively set to zero, obviating the need to accumulate information from the ab initio point involved. Most applications of fitted PESs systematically sample the PES. Software techniques can then be used to subdivide a large set of ab initio points into periodically updated collections near and far from the region of the PES currently being probed. Then only the collection of near ab initio points needs to be investigated at all for an IMLS fit. In other words, the nature of the weights offer opportunities to reduce the global reach of an evaluation of an IMLS fit into a more local reach.

The results of the previous section do not answer the question of what is the optimal degree for an IMLS fit. Two contending factors are involved that must be evaluated on a case-by-case basis. The first factor is that the time to evaluate an IMLS fit goes as the square of the degree. The second factor is that a higher-degree IMLS fit requires fewer ab initio points to achieve a given level of accuracy. How pronounced this second factor is depends on the manner of selecting ab initio points and the overall

accuracy required. Fine grids and high accuracy can display large changes in the value of N needed for a given m (see Fig. 7). Automatic point selection and less demanding accuracy can make adjacent values of m have essentially the same values of N (see Table 1). Consequently, the optimal degree of IMLS cannot be answered in any general way. However, all the results of this work indicate that there are more advantages to higher-degree IMLS methods than to the Shepard method. In practice today, the Shepard method is applied only in a modified form involving separate fits to value, gradients, and Hessians. The comparison of higher-degree IMLS fits to just values compared to modified Shepard fits to values, gradients, and Hessians is beyond the scope of this paper but will be the subject of a future study. Since the origin of the modification to Shepard is due to poor derivative properties and because Hessians are expensive to compute, one would suspect that higher-degree IMLS methods will compete well in overall cost with modified Shepard methods. Alternatively, just like Shepard, higher-degree IMLS methods can be modified, perhaps to beneficial effect.

The results of the previous section for the MO test case allow some general statements regarding splines and IMLS. The results show that TD-IMLS is considerably superior in accuracy to cubic splines, even though both methods are polynomial fits of the same degree. As discussed earlier, the weights in IMLS methods give the nominal polynomial fit additional nonlinear flexibility. While splines can be defined with higher degrees than cubic, the results in this paper would suggest that the comparable higher-degree IMLS method will always be superior. This superiority does come at a cost. In multidimensional applications, the ab initio points used in a spline can be considered as the vertices of hypercubes in internal degree of freedom space. Within each hypercube, the

coefficients of the spline polynomial can be computed once and for all and saved. Then, the evaluation of the spline is no more complicated than a simple retrieval of the coefficients and an evaluation of the polynomial. In contrast, the IMLS method must evaluate from scratch the polynomial coefficients and then use them, as in the spline, to evaluate the polynomial. As discussed above, the coefficient evaluation may be relatively easily restricted to local ab initio information while the once-and-for-all evaluation of the spline coefficients is inherently global. Thus the trade-offs between spline and higher-degree IMLS can be complicated. However, comparable-degree IMLS methods will require fewer ab initio points for a given accuracy and are easier to program.

Reproducing kernel Hilbert space methods are generalizations of splines that depart from simple polynomial descriptions. Many unweighted least squares fitting methods similarly depart from simple polynomial descriptions (e.g., many-body fitting methods or rotating Morse oscillator methods). It is beyond the scope of this paper to systematically compare higher-degree IMLS methods with the great variety of fitting approaches that have been used. We note, however, that many apparently different fitting methods employ common strategies that can equally well be applied to IMLS methods. For example, polynomial fits in x can be regarded as expansion in a basis set of x^0 , x^1 , x^2 , etc. The fitting approach does not essentially change if one changes the basis set to something completely different, such as 1 , e^{-x} , e^{-2x} , etc., as would apply in a rotating Morse oscillator approach. Nor does the fitting approach essential change if one transforms the coordinate, for example, $x \rightarrow 1/(x - x_e)$ as in Dunham expansions. Both

basis set changes and coordinate transformations can be easily accommodated within a higher-degree IMLS framework.

V. Conclusion

We have presented the basic formal and numerical aspects of higher-degree IMLS methods in the context of 1-D applications for both the potential and its derivatives using two relatively different test cases. For these applications we have systematically examined the effect of weight function parameters, the degree of the IMLS fit, and the number and placement of ab initio points. From this systematic behavior, we have discovered regions of parameter space for the weight functions that allow compact and accurate representations of potentials and their derivatives for first-degree and higher-degree IMLS fits. We have documented how the number of ab initio points needed to achieve a given accuracy declines with the degree of the IMLS. We have outlined automatic procedures for ab initio point selection that can optimize this decline.

The results of this systematic study and our earlier study on one particular surface support further studies of IMLS methods. Such studies are in progress for a variety of higher dimensional potential energy surfaces. Other studies are being planned for the incorporation of IMLS techniques into trajectory routines. In addition, the direct incorporation of gradients and Hessians into the IMLS framework is being explored.

Acknowledgments: This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, and the Mathematical,

Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy under Contract No. W-31-109-Eng-38 (Argonne) and Contract No. DE-FG02-01ER15231 (OSU).

References

- ¹ M. E. Tuckerman and G. J. Martyna, *J. Phys. Chem. B*, **104**, 159 (2000).
- ² R. J. Duchovic, Yu. L. Volobuev, G. C. Lynch, D. G. Truhlar, T. C. Allison, A. F. Wagner, B.C. Garrett, J. C. Corchado, *Comp. Phys. Comm.* **144**, 169 (2002).
- ³ G. C. Schatz, *Rev. Mod. Phys.* **61**, 669 (1989).
- ⁴ J. N. Murrell, S. Carter, S. C. Farantos, P. Huxlay, and A. Varandas, *Molecular Potential Energy Functions* (Wiley, Chichester, 1984).
- ⁵ G. K. Chawla, G. C. McBane, P. L. Houston, and G. C. Schatz, *J. Chem. Phys.* **88**, 5481 (1988).
- ⁶ G. C. Schatz, A. Papaioannou, L. A. Pederson, L. B. Harding, T. Hollebeek, T. S. Ho, and H. Rabitz, *J. Chem. Phys.* **107**, 2340 (1997).
- ⁷ H. Koizumi, G. C. Schatz, and S. P. Walch, *J. Chem. Phys.* **95**, 4130 (1991).
- ⁸ A. J. Dobbyn, J. N. L. Connor, N. A. Besley, P. J. Knowles, and G. C. Schatz, *Phys. Chem. Chem. Phys.* **1**, 957 (1999).
- ⁹ J. Ischtwan and M. A. Collins, *J. Chem. Phys.* **100**, 8080 (1994).
- ¹⁰ R. P. A. Bettens and M. A. Collins, *J. Chem. Phys.* **111**, 816 (1999).
- ¹¹ M. Yang, D. H. Zhang, M. A. Collins, and S.-Y. Lee, *J. Chem. Phys.* **114**, 4759 (2001).
- ¹² R. Farwig, in *Algorithms for Approximation*, edited by J. C. Mason and M. G. Cox (Clarendon, Oxford, 1987).
- ¹³ D. H. McLain, *Comput. J.* **17**, 318 (1974).
- ¹⁴ T. Ishida and G. C. Schatz, *Chem. Phys. Lett.* **314**, 369 (1999).
- ¹⁵ G. C. Schatz, in *Reaction and Molecular Dynamics, Lecture Notes in Chemistry, Vol. 14*, edited by A. Lagana and A. Riganelli (Springer, Berlin, 2000), p. 15.
- ¹⁶ G. G. Maisuradze and D. L. Thompson, *J. Phys. Chem. A* (in press).
- ¹⁷ In the work by Koizumi et al. the potential as a function of r was fitted by a Morse function. However, at most values of R and θ (angle between vectors pointing along R and r) there were not enough r points available to fit all Morse parameters, so some of the parameters were fixed for certain ranges of R (Morse parameter D_e was frozen for all R and θ). The effect of N-N motion was introduced by an additive correction that was supported by only a relatively limited number of ab initio points. The result in the 1-D slice we chose was a small amplitude bump both before and after the major peak. There is no physical reason for such bumps; they are just a result of the sparse set of ab initio points and are of a scale smaller than the level of accuracy Koizumi et al. were interested in. Since we are examining high degrees and highly accurate fits in the presented paper, we found it convenient to remove the correction term that Koizumi et al. devised.
- ¹⁸ P. Lancaster and K. Salkauskas, *Curve and Surface Fitting, An Introduction* (Academic, London, 1986).

-
- ¹⁹ J. R. Rice, *The Approximation of Functions, Vol. II—Advanced Topics* (Addison-Wesley, Reading, 1969), p. 161.
- ²⁰ K. C. Thompson and M. A. Collins, *J. Chem. Soc., Faraday Trans.* **93**, 871 (1997).
- ²¹ D. F. R. Brown, M. N. Gibbs, and D. C. Clary, *J. Chem. Phys.* **105**, 7597 (1996).
- ²² A. F. Wagner, G. C. Schatz, and J. M. Bowman, *J. Chem. Phys.* **74**, 4960 (1981)

Figure Captions

- Fig. 1: The potential in kcal/mol for both the MO (dashed line) and NH₂ (solid line) test cases over the range of R in a_0 used in the fitting procedures.
- Fig. 2: The rms error for the potential energy in kcal/mol versus ϵ for Shepard (dotted line), FD-IMLS (dash dotted line), and SD-IMLS (solid line). Both MO (upper lines) and NH₂ (bottom lines) test cases are displayed. In these calculations, $n = 6$, $N = 33$, and the ab initio points are equally spaced.
- Fig. 3: As in Fig. 2 only for $N = 17$.
- Fig. 4: As in Fig. 2 only for the rms error of the first derivative in kcal/mol/ a_0 .
- Fig. 5: For the NH₂ test case, the first derivative in kcal/mol/ a_0 versus R in a_0 for (a) Shepard [solid line ($\epsilon = 1 \times 10^{-13}$), dash dotted line ($\epsilon = 1 \times 10^{-7}$), dashed line ($\epsilon = 1 \times 10^{-4}$), and the exact potential (dotted line)]; (b) FD-IMLS [solid line ($\epsilon = 1 \times 10^{-13}$), dash dotted line ($\epsilon = 1 \times 10^{-7}$), dashed line ($\epsilon = 1 \times 10^{-4}$), and the exact potential (dotted line)]. In both cases $n = 6$ and $N = 33$.
- Fig. 6: As in Fig. 2 only for rms error versus n for optimal ϵ . See text for details.
- Fig. 7: The rms error for the potential energy in kcal/mol versus N for equally spaced points for both HN₂ (a) and MO (b) test cases. The solid line with circles denotes cubic spline. Seven different-degree IMLS fits are denoted by Shepard ($\bullet \bullet \bullet$), FD-IMLS (- - -), SD-IMLS (—), fourth-degree IMLS (— $\bullet \bullet$ —), seventh-degree IMLS (— — —), and ninth-degree IMLS (- \bullet -). The parameter $n = 6$ and optimal values of ϵ are used (see text for details).
- Fig. 8: As in Fig. 7 only for the rms error for the first derivative.

Fig. 9: The difference between IMLS fits and the exact potential in kcal/mol versus R in a_0 for HN_2 (a) and MO (b) test cases. The lines denote Shepard (dotted), FD-IMLS (dashed), SD-IMLS (dashed dotted) and TD-IMLS (solid). $N=17$ and $n = 6$ for optimal values of ϵ .

Fig. 10: For two different values of N , the calculated value of ΔR as a function of the degree of the IMLS fit. Four different options for calculating ΔR are indicated: by absolute difference between IMLS fit and exact MO potential, (—); by absolute difference between IMLS fits of degree m and $(m-1)$, (- - -); by 1.5 times the absolute difference between IMLS fits of degree m and $(m-1)$, (—••—); and by 3.0 times the absolute difference between IMLS fits of degree m and $(m-1)$, (— — —). The upper set of four curves are for $N=87$. The bottom set are for $N=16$ where the plotted value is $0.1 \times \Delta R$ so as to better space the curves in the figure. See text for details.

Table I. Number of data points needed to achieve the certain rms errors with different point selection strategies using FD-, SD-, TD-, and FRD-IMLS for HN₂ and MO.

IMLS Degree	HN ₂			MO		
	rms error (kcal/mol)			rms error (kcal/mol)		
	4.00	1.00	0.25	4.00	1.00	0.25
FD						
FD-ZD	14	21	47	8	20	45
FD-exact	6	13	23	8	13	24
FD (equal grid)	9	17	33	17	33	65
SD						
SD-FD	8	9	20	7	12	27
SD-exact	7	8	17	7	10	17
SD (equal grid)	9	9	33	17	33	65
TD						
TD-SD	9	11	15	7	9	14
TD-exact	7	8	17	7	9	13
TD (equal grid)	9	9	33	17	17	33
FRD						
FRD-TD	9	9	17	7	10	12
FRD-exact	8	9	16	7	9	13
FRD (equal grid)	9	9	17	9	17	33

Table II. Number of data points needed to achieve the certain rms errors using TD-
IMLS for HN₂ and MO.

Rms Error (kcal/mol)	TD-SD			FRD-TD		
	5 Seed Pts. Adding 1 Pt	6 Seed Pts Adding 1 Pt	5 Seed Pts Adding 2 Pts	5 Seed Pts. Adding 1 Pt	6 Seed Pts Adding 1 Pt	5 Seed Pts Adding 2 Pts
HN ₂						
4.00	9	9	11	9	8	9
1.00	11	14	11	9	10	11
0.25	15	19	15	17	17	17
MO						
4.00	7	7	7	7	7	9
1.00	9	9	9	10	9	11
0.25	14	15	13	12	11	13

Table III. Number of data points needed to achieve the certain rms errors using FRD-
IMLS for HN₂ and MO.

Rms Error (kcal/mol)	TD			FRD		
	Rms-Difference/Rms-Error			Rms-Difference/Rms-Error		
	2.0	1.0	0.5	2.0	1.0	0.5
HN ₂						
4.00	5 (7.98)	5 (7.98)	6 (5.45)	5 (12.23)	7 (5.47)	7 (5.47)
1.00	8 (4.16)	9 (1.31)	11 (0.83)	7 (5.47)	9 (0.77)	11 (0.66)
0.25	11 (0.83)	14 (0.32)	15 (0.24)	11 (0.66)	12 (0.49)	14 (0.38)
MO						
4.00	5 (17.84)	6 (4.63)	7 (3.42)	5 (15.31)	6 (5.06)	7 (2.37)
1.00	8 (1.70)	10 (0.52)	13 (0.25)	7 (2.37)	9 (1.18)	11 (0.26)
0.25	13 (0.25)	16 (0.10)	16 (0.10)	11 (0.26)	12 (0.22)	15 (0.20)

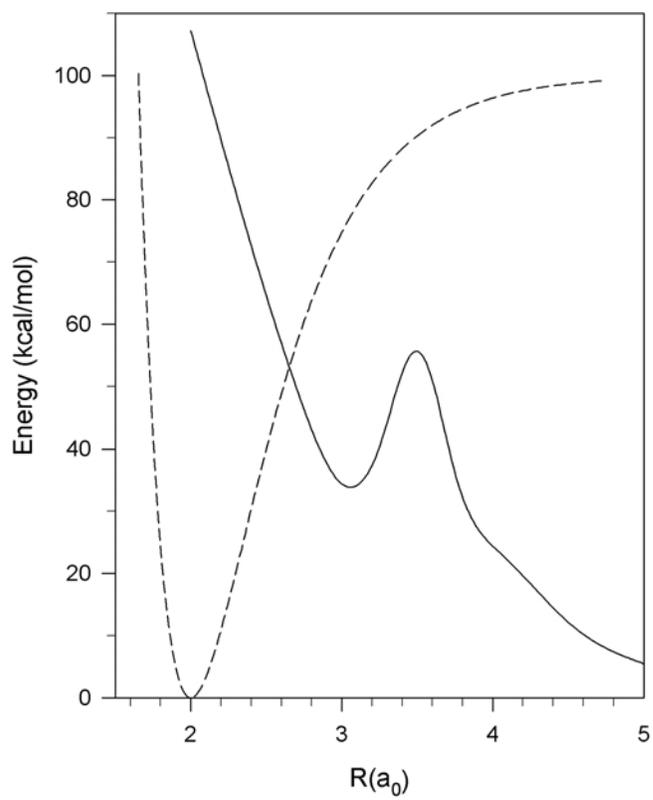


Fig. 1

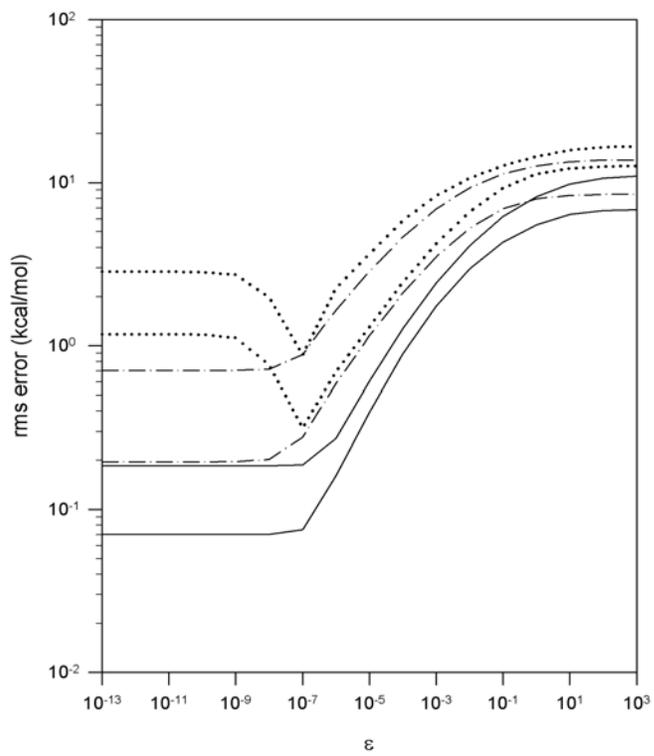


Fig. 2

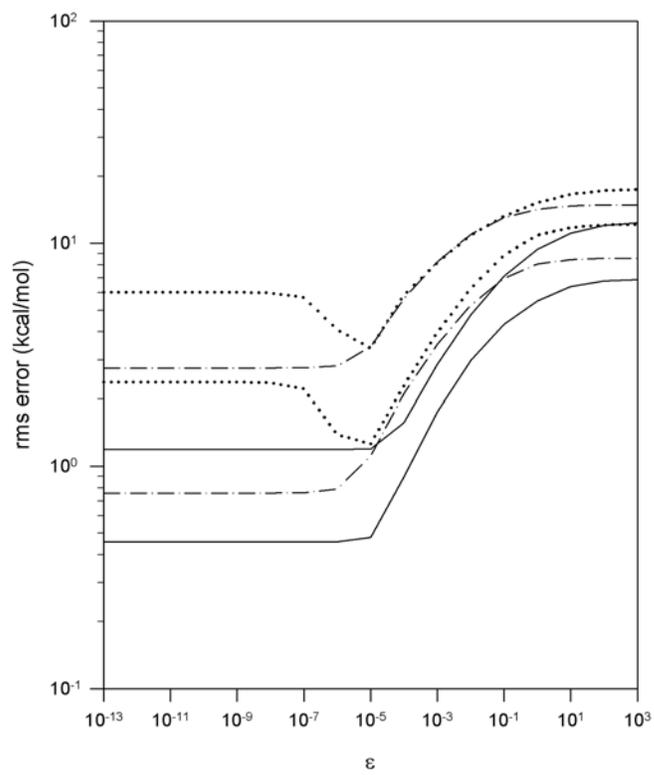


Fig. 3

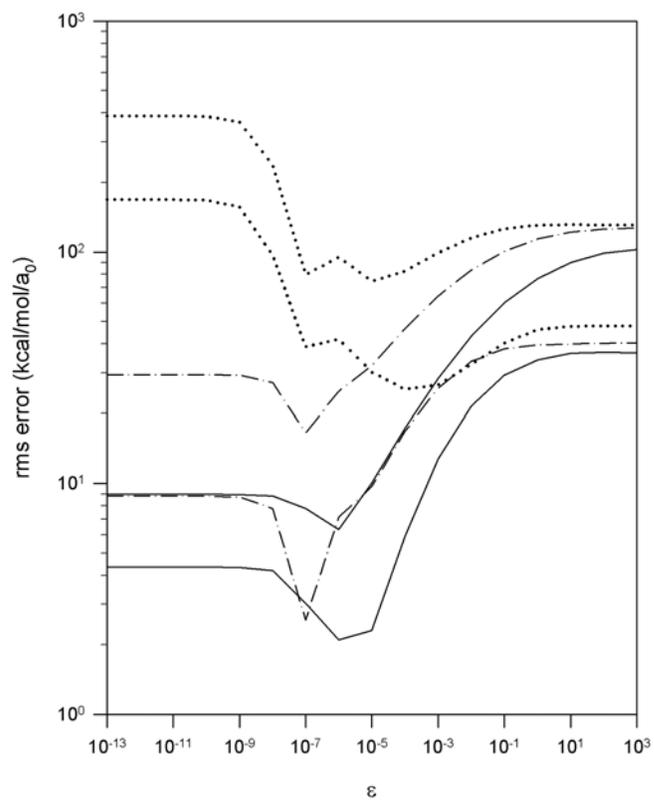


Fig. 4

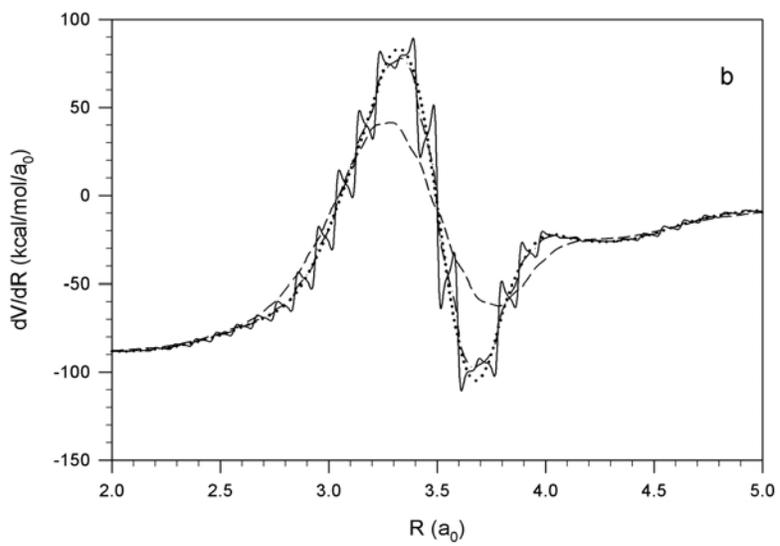
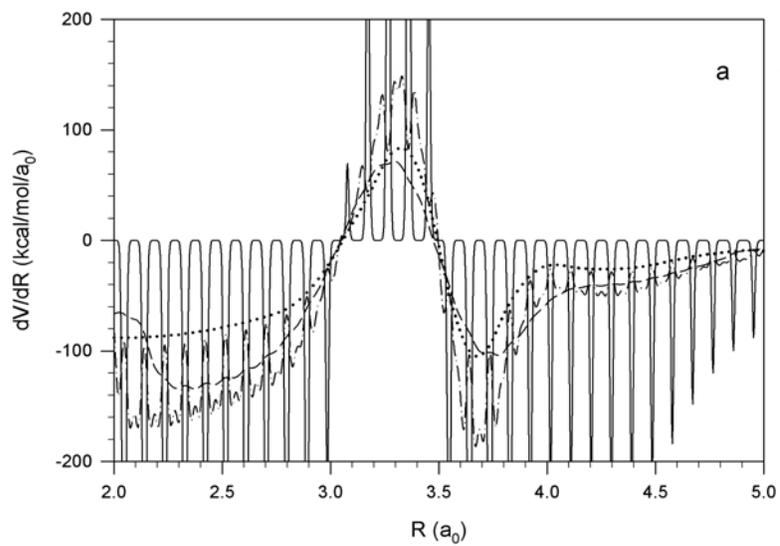


Fig. 5

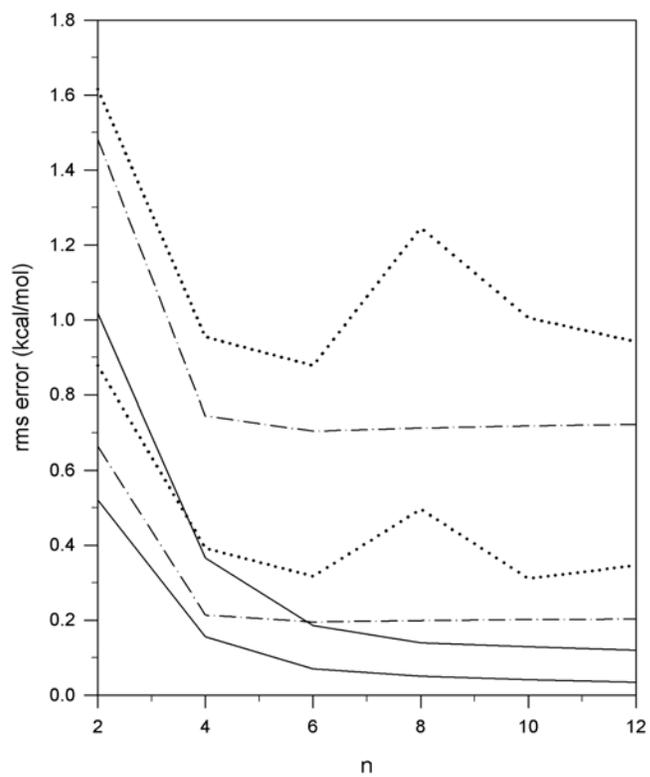


Fig. 6

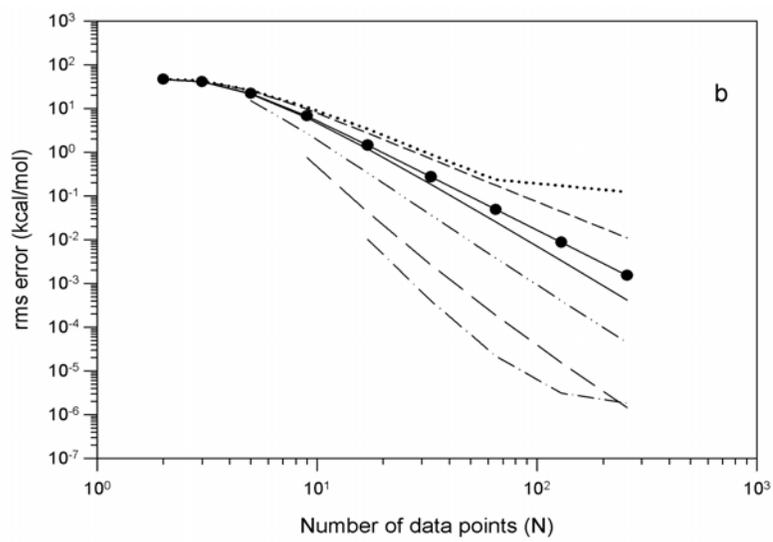
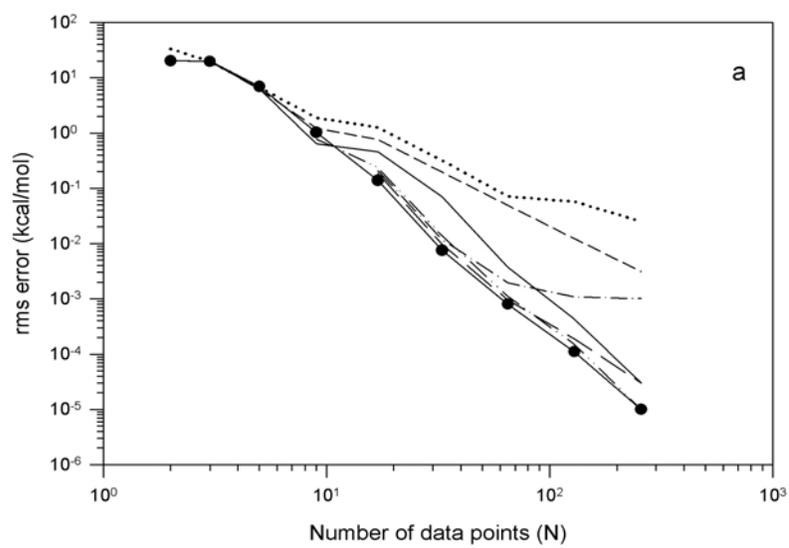


Fig. 7

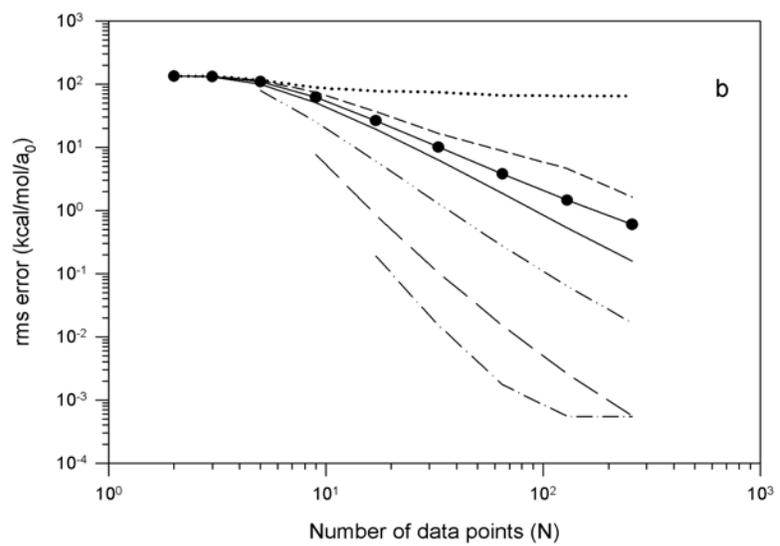
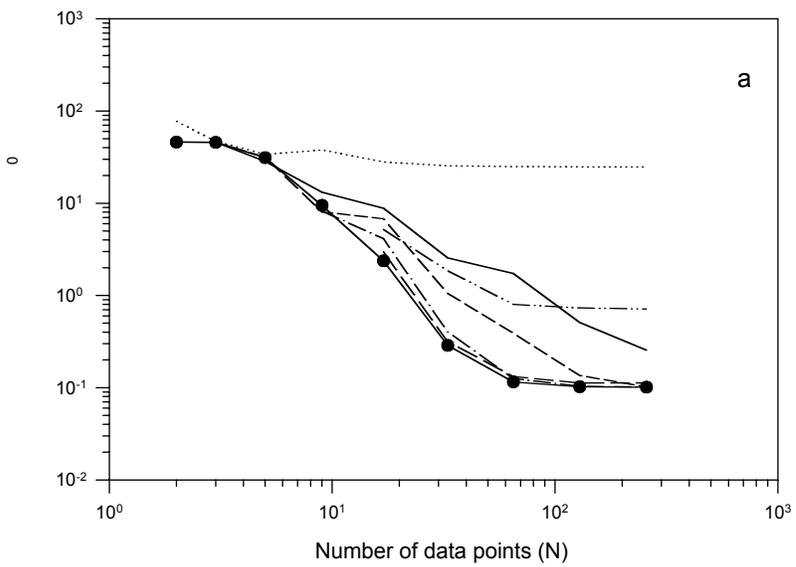


Fig. 8

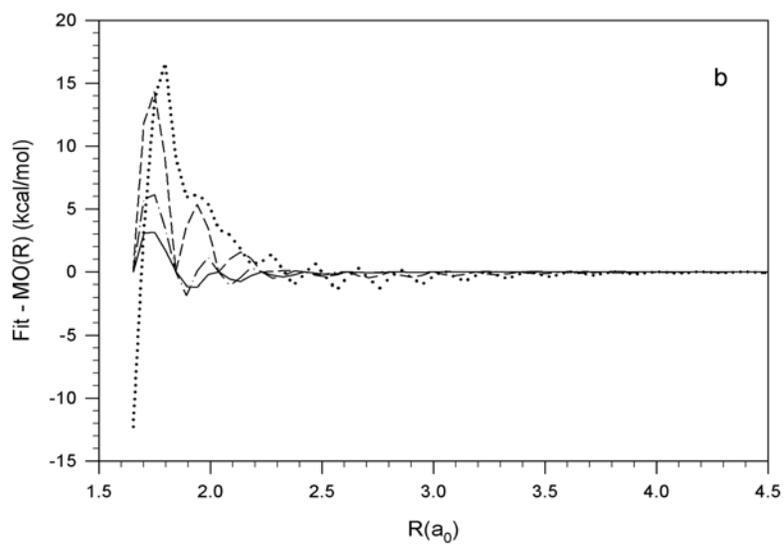
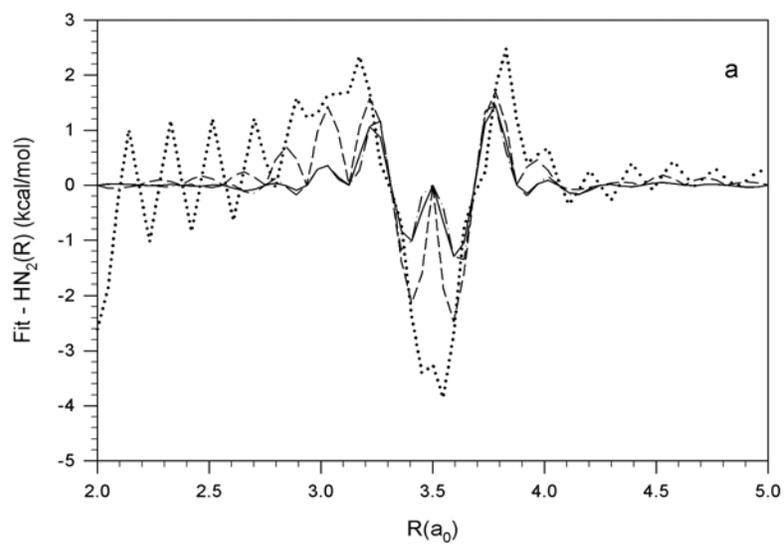


Fig. 9

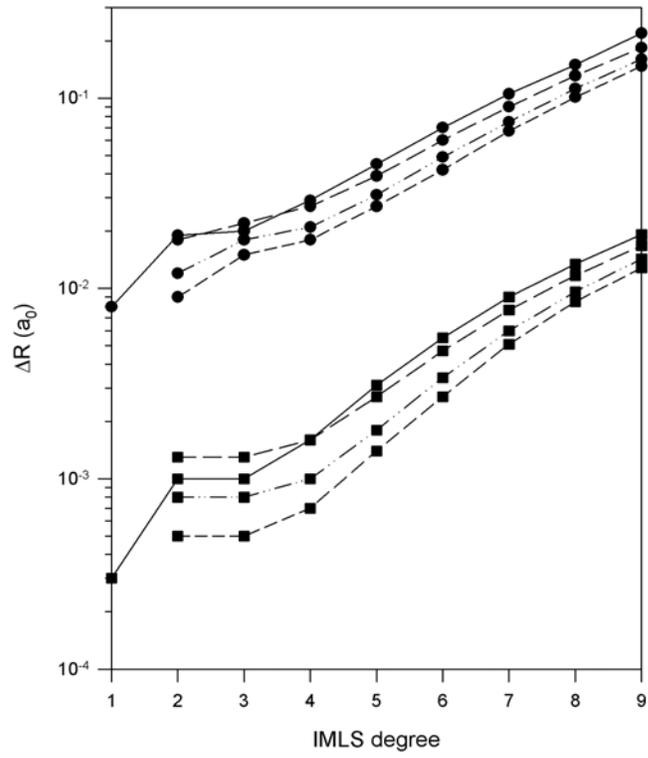


Fig. 10