Theor Chem Account (2007) 118:755–767
DOI 10.1007/s00214-007-0358-7

REGULAR ARTICLE

Gradient incorporation in one-dimensional applications of interpolating moving least-squares methods for fitting potential energy surfaces

Igor V. Tokmakov · Albert F. Wagner · Michael Minkoff · Donald L. Thompson

Received: 9 February 2007 / Accepted: 26 March 2007 / Published online: 21 June 2007

Abstract We present several approaches to use gradients in higher degree interpolating moving least squares (IMLS) methods for representing a potential energy surface (PES). General procedures are developed to obtain smooth approximations of the PES and its derivatives from quasi-uniform sets of energy and gradient data points. These methods are illustrated and analyzed for the Morse oscillator and a 1-D slice of the ground-state PES for the HCO radical computed using density functional theory. Variations in the IMLS fits with the number and distribution of points and the degree of the polynomial fitting basis set are examined. We determine the effects of gradient inclusion on the accuracy of the IMLS values of the energy, first and second derivatives for two 1-D test cases. Gradient inclusion reduces the number of data points required by up to 40%.

Keywords Interpolating moving least squares · Potential energy surface · Polynomial fitting

1 Introduction

An accurate potential energy surface (PES) is essential for theoretical studies of molecular and reaction dynamics. The option of using high-level ab initio methods has grown in recent years because of the improved accuracy of quantum chemistry methods and the availability of greater computing capabilities. This is especially significant for studying chemical reactions because high-level electronic structure theory is often required for accurate descriptions of bond breaking and formation. We are interested in developing efficient (i.e., requiring a minimum of ab initio calculations), general, and robust numerical fitting procedures for ab initio PESs that can be automated to provide values of the energy, gradients, and hessians with controllable accuracy.

The local PES fitting method introduced by Ischtwan and Collins [1], which is based on modified Shepard interpolation [2–6], is such a method. The *unmodified* Shepard method [7,8] approximates the PES at a given evaluation point as a weighted average of the exact values at the neighboring data points. Larger weights are assigned to the data points closer to the evaluation point, making the interpolation local. The Shepard method is a simple and general method of fitting multivariate surfaces from scattered data points, but it suffers from the unphysical flat-spot phenomenon whereby the derivatives of the interpolated surface are zero at every data point [8]. The modified Shepard method cures this problem by fitting the local Taylor expansions about the data points instead of values alone. The Taylor expansions must at least include second-order terms to avoid the flat-spot phenomenon in the fitted potential and its gradient. Collins and co-workers [9–16] have studied various aspects of fitting PESs by the modified Shepard interpolation method and have developed automated PES-growing procedures based on different selection criteria, such as uncertainty of the interpolated energy [16] and importance sampling, whereby new data points are placed iteratively in the regions of configuration space important for the properties of interest. In particular, trajectory sampling has been used for reactive systems to place data

I. V. Tokmakov
Department of Chemistry, University of Missouri, Columbia, MO 65211, USA

A. F. Wagner · M. Minkoff
Chemistry Division, Argonne National Laboratory, Argonne, IL 60439, USA

D. L. Thompson (✉)
Department of Chemistry, University of Missouri, Columbia, MO 65211, USA
e-mail: thompsondon@missouri.edu

points in the dynamically important regions of configuration space [1]; approximate vibrational energy levels and associated wavefunctions have been calculated to bias the fit for bound systems [13]; and quantum sampling regimes have been developed for quantum diffusion Monte Carlo simulations [15].

Attractive features of the modified Shepard approach are its mathematical simplicity and ease of automation, but the need of the second- or higher order derivatives cannot be readily or inexpensively satisfied by high-level ab initio calculations. To address this problem Ishida and Schatz [17, 18] suggested obtaining approximate first- and second-order derivatives by the interpolating moving least squares (IMLS) method [8,19–26] and using them in the modified Shepard fit. Recently, we have studied the use of IMLS in more detail for fitting PESs [27–32]. The IMLS method involves a basis of polynomials up to any desired degree k. (The Shepard method is a zeroth-degree IMLS method.) Surfaces fit by second- and higher degree IMLS have well-behaved first and second derivatives, i.e., there is no flat-spot problem and gradients and Hessians are not required. This is an important advantage for fitting PESs obtained by high-level ab initio calculations.

Our earlier work focused on the features of the IMLS fits of 1-D [28] and 3-D PESs [27]. These cases highlight the improved accuracy in fitted values and derivatives obtained with higher degree IMLS. In order to improve the accuracy and efficiency of the numerical fitting methods, we have recently used a dual-level approach in which we fit the difference of the exact and a reference surface [32]. This approach was used on a 6-D PES of HOOH with two interpolation methods: modified Shepard and second-degree (2d) IMLS. The results demonstrated that with the dual-level approach the 2d-IMLS and modified Shepard methods are comparably accurate in fitting the same number of ab initio points. However, the IMLS requires only the energy values, and not the gradients and Hessians [32]. We have also developed several strategies for selectively eliminating distant points, which have little or no effect on the fitted PES, that make the fitting much more efficient [31].

Although the IMLS methods do not require derivatives at the data points, they can be used in the fitting. Many quantum chemistry methods can compute gradients analytically at an additional cost of 10–100% of the energy calculation [33–36]. Available gradients can be used to improve the fitting accuracy, thus requiring fewer expensive quantum chemistry calculations needed to fit a PES to a specified accuracy. The present study explores several ways of fitting energy and gradient data using IMLS. We have investigated the benefits of gradient incorporation for two representative 1-D potentials: the Morse oscillator (V_{MO}) and a reactive profile (V_{HCO}) for constrained dissociation of the formyl radical to H and CO. In earlier studies [27,28] we have examined the quality of IMLS fits for the same V_{MO} potential and for a collinear cut through the analytical HN_2 PES of Koizumi et al. [37]. However, the HN_2 PES is a fit by cubic splines [38]. The inherent cubic nature of the fit makes the application of higher-than-cubic IMLS fits problematic, as the results of our previous 1-D study indicated [28]. Thus, in the present study we used a non-collinear cut through the HCO PES obtained directly from electronic structure calculations.

The paper is arranged as follows. Section 2 briefly reviews the standard IMLS method. Various approaches for incorporating gradient data in the IMLS fitting procedure are presented in Sect. 3. The weights, sampling, and model potentials are described in Sect. 4. In Sect. 5 we describe the results of the application of several approaches to fitting potential and gradient data and evaluate the benefits of gradient incorporation in the IMLS. A summary and conclusions are given in Sect. 6.

2 IMLS methods without derivatives

Let an arbitrary geometry of a polyatomic system with d internal coordinates be specified by a vector $\mathbf{z}$. We seek an approximation of a multidimensional PES from a set of ab initio energies evaluated at reference geometries, which will be called data points. An approximate PES representation is given by a linear combination of m basis functions:

$$V_{\mathrm{fit}}(\mathbf{z}) = \sum_{j=1}^{m} a_j(\mathbf{z}) b_j(\mathbf{z}) = \mathbf{b}^{\mathrm{T}}(\mathbf{z})\ \mathbf{a}(\mathbf{z}), \tag{1}$$

where $\mathbf{b}(\mathbf{z}) = (b_1(\mathbf{z}), b_2(\mathbf{z}), \ldots, b_m(\mathbf{z}))^{\mathrm{T}}$ is the vector of linearly independent basis functions, and $\mathbf{a}(\mathbf{z}) = (a_1(\mathbf{z}), a_2(\mathbf{z}), \ldots, a_m(\mathbf{z}))^{\mathrm{T}}$ is the vector of expansion coefficients. Given accurate potential energy values $\{V(\mathbf{z}^{(i)})\}$ for a set of n data points: $\{\mathbf{z}^{(i)}, i = 1, 2, \ldots, n\}$, the expansion coefficients are determined by minimizing the weighted least-squares error functional:

$$E_0(\mathbf{z}) = \sum_{i=1}^{n} w(\|\mathbf{z} - \mathbf{z}^{(i)}\|)[V(\mathbf{z}^{(i)}) - \mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\ \mathbf{a}(\mathbf{z})]^2 \tag{2}$$

where $w(r)$ is a non-negative weight function such that $w(r \to \infty) \to 0$. *Interpolative* weights have a singularity at the origin: $w(r \to 0) \to \infty$, which ensures that IMLS fits reproduce the PES exactly at the data points. In practice, this singularity can be avoided [$w(r \to 0) \to 1/\varepsilon$] at some negligible cost in interpolative rigor if ε is small enough. In addition to these properties, both weight and basis functions should be sufficiently smooth to ensure smooth behavior of the interpolant $V_{\mathrm{fit}}(\mathbf{z})$.

The minimization conditions $\{\partial E_0/\partial a_j = 0\}$ lead to the IMLS normal equation, which can be written in a compact matrix–vector notation [8]:

$$\mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{B}\mathbf{a}(\mathbf{z}) = \mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{f}, \tag{3}$$

where $\mathbf{B}$ is the $n \times m$ matrix of the basis function values at the data points,

$$\mathbf{B} = \begin{pmatrix} b_1(\mathbf{z}^{(1)}) & b_2(\mathbf{z}^{(1)}) & \dots & b_m(\mathbf{z}^{(1)}) \\ b_1(\mathbf{z}^{(2)}) & b_2(\mathbf{z}^{(2)}) & \cdots & b_m(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(\mathbf{z}^{(n)}) & b_2(\mathbf{z}^{(n)}) & \cdots & b_m(\mathbf{z}^{(n)}) \end{pmatrix}, \tag{4}$$

$\mathbf{W}(\mathbf{z})$ is the diagonal positive definite weight matrix:

$$\mathbf{W}(\mathbf{z}) = \mathrm{diag}(w(\|\mathbf{z} - \mathbf{z}^{(1)}\|),\ w(\|\mathbf{z} - \mathbf{z}^{(2)}\|), \\ \dots,\ w(\|\mathbf{z} - \mathbf{z}^{(n)}\|)), \tag{5}$$

and $\mathbf{f}$ is the column vector of the potential energy values at the data points:

$$\mathbf{f} = (V(\mathbf{z}^{(1)}),\ V(\mathbf{z}^{(2)}),\ \dots,\ V(\mathbf{z}^{(n)}))^{\mathrm{T}}. \tag{6}$$

Provided the number and density of data points are sufficient to specify the fit, the symmetric $m \times m$ matrix $\mathbf{S}_0(\mathbf{z}) = \mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{B}$ has full rank [Rank $(\mathbf{S}_0(\mathbf{z})) = m$] and the solution of Eq. (3) is:

$$\mathbf{a}(\mathbf{z}) = \mathbf{S}_0^{-1}(\mathbf{z})\mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{f}. \tag{7}$$

In the case of the zero-degree IMLS, the solution is very simple and is given by the Shepard formula:

$$V_{\mathrm{fit}}(\mathbf{z}) = a_1(\mathbf{z}) = \frac{\sum_{i=1}^{n} w(\|\mathbf{z} - \mathbf{z}^{(i)}\|)V(\mathbf{z}^{(i)})}{\sum_{i=1}^{n} w(\|\mathbf{z} - \mathbf{z}^{(i)}\|)}. \tag{8}$$

This simplicity carries over to the modified Shepard method, which uses the same formula for the fitted PES but the $V(\mathbf{z}^{(i)})$ values are replaced with the estimate of $V(\mathbf{z})$ by the local Taylor expansions about $\mathbf{z}^{(i)}$. In this study, we have examined the usual quadratic modified Shepard method (hereafter denoted m-Shepard) in comparison with IMLS fits of different degrees.

Higher degree IMLS fits require an inverse of the $\mathbf{S}_0(\mathbf{z})$ matrix to be evaluated at every point. In practice, a direct inversion of matrix $\mathbf{S}_0(\mathbf{z})$ can lead to severely ill-conditioned numerical problems; thus, the normal equation is solved by QR or singular value decomposition (SVD) techniques. These methods have several advantages including treating the rank-deficient case and improved numerical conditioning.

Chemical applications often require approximate first- and second-order derivatives of the PES. Differentiation of Eq. ((1)) leads to the following expressions for the first derivatives of the fitted potential ($\partial_k = \partial/\partial z_k, k = 1, 2, \dots, d$):

$$\partial_k V_{\mathrm{fit}}(\mathbf{z}) = \partial_k \mathbf{b}^{\mathrm{T}}(\mathbf{z})\mathbf{a}(\mathbf{z}) + \mathbf{b}^{\mathrm{T}}(\mathbf{z})\partial_k \mathbf{a}(\mathbf{z}). \tag{9}$$

The derivatives of the IMLS coefficients can be derived by differentiating Eq. (3):

$$\mathbf{S}_0(\mathbf{z})\partial_k \mathbf{a}(\mathbf{z}) = \mathbf{B}^{\mathrm{T}}(\partial_k \mathbf{W}(\mathbf{z}))[\mathbf{f} - \mathbf{B}\mathbf{a}(\mathbf{z})]. \tag{10}$$

Since Eqs. (3) and (10) have the same matrix $\mathbf{S}_0(\mathbf{z}) = \mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{B}$ on the left-hand side, SVD or QR decompositions performed to solve Eq. (3) can be reused in the calculation of $\partial_k \mathbf{a}(\mathbf{z})$. Similarly, differentiating Eq. (10) leads to an expression for the second derivative that reuses the matrix decomposition.

While the derivatives of IMLS fitted functions can be exactly calculated at an additional cost of evaluating the derivatives of IMLS coefficients, others such as Farwig [5] and Levin [25] have recommended the computationally simpler one-term derivative approximation:

$$\partial_k V_{\mathrm{fit}}(\mathbf{z}) \approx \partial_k \mathbf{b}^{\mathrm{T}}(\mathbf{z})\ \mathbf{a}(\mathbf{z}), \tag{11}$$

which ignores the second term in Eq. (9). If the constant term is included in the basis (e.g. $b_1(\mathbf{z}) = 1$) and the interpolative weight function is used, then vector $\partial_k \mathbf{a}(\mathbf{z}^{(i)})$ is orthogonal to the vector of basis functions at the data points (see Appendix). By continuity, the missing $(\mathbf{b}^{\mathrm{T}}\partial_k \mathbf{a})$ term in Eq. (9) is negligibly small in the local neighborhood of data points. The "one-term" derivative approximation then just assumes that it can be ignored globally. This approximation can be systematically extended to the higher order derivatives. For example, approximate second derivatives are expressed as:

$$\partial_j \partial_k V_{\mathrm{fit}}(\mathbf{z}) \approx (\partial_j \partial_k \mathbf{b}^{\mathrm{T}}(\mathbf{z}))\ \mathbf{a}(\mathbf{z}). \tag{12}$$

Using the sample 1-D PES fits described in Sect. 4, we will show that Eqs. (11) and (12) do in fact provide good approximations of the first and second derivatives, respectively.

3 IMLS methods with incorporation of derivatives

As in the standard procedure, we define the fit in the form of Eq. (1) but, in addition to minimizing the errors in the fitted function [Eq. (2)], we also want to minimize the weighted errors in each of its gradient components ($k = 1, 2, \dots, d$):

$$E_k(\mathbf{z}) = \sum_{i=1}^{n} w_k(\|\mathbf{z} - \mathbf{z}^{(i)}\|)[\partial_k V(\mathbf{z}^{(i)}) - \partial_k \mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\ \mathbf{a}(\mathbf{z})]^2. \tag{13}$$

In general, different weight functions can be used for different gradient components. However, here we will consider the same functional form $w_k(r) = w(r)$ for simplicity.

In order to fit the PES and its gradient simultaneously, the IMLS coefficients are determined by minimizing the combined error functional $E(\mathbf{z})$ that incorporates both energy and gradient data:

$$E(\mathbf{z}) = \sum_{k=0}^{d} \lambda_k^2(\mathbf{z}) E_k(\mathbf{z}), \tag{14}$$

where $\lambda_k(\mathbf{z})$'s are positive factors that scale the weighted errors for individual surfaces and make each term in the combined error functional independent of physical units. Multiplying Eqs. (2) and (13) by scaling factors does not affect the minimization of individual error functionals, but λ_k's control *relative* fitting accuracy for different surfaces in multi-surface fitting. Depending on the fitting preferences, the choice of scaling factors is not unique. In chemical applications, accurate representations of the PES and its derivatives are typically required. Also, available ab initio data (both energies and gradients) are of high numerical precision. Taking this into account we put two constraints on the selection of λ_k's. The first is that at each data point, the fit should recover both the value and the derivatives at the data points. Once the first constraint is met, the second constraint is that to the degree possible, λ_k's must be selected to minimize fitting error.

Concerning the first constraint, the IMLS basis always includes a constant basis function, i.e., $b_1(z) = 1$, whose derivative is zero. The coefficient of that basis function has no effect on the functionals in Eq. (13) but does affect the E_0 functional of Eq. (2). As such, that coefficient alone is determined independently of scaling factors in a way that ensures that the IMLS fit reproduces energy data values. If $\lambda_k/\lambda_0 (k = 1, 2, \ldots, d)$ is then large enough so that the functionals of Eq. (13) dominate the combined functional, then it is straightforward to show that the derivative data will be recovered by the fit. Thus the effect of the first constraint is to require $\lambda_k/\lambda_0 (k = 1, 2, \ldots, d)$ be large enough to reproduce gradients at the data points.

With regard to the second constraint of minimizing the fitting error, there is no simple and general implication for λ_k's that can be derived. Having tried several choices, we selected one of the simpler definitions of λ_k's in terms of the dynamic ranges for the potential and its derivatives estimated from the data points used to define the fit:

$$\begin{aligned} \lambda_0^{-1} &= \max\{V(\mathbf{z}^{(i)})\} - \min\{V(\mathbf{z}^{(i)})\}, \\ \lambda_k^{-1} &= \max\{\partial_k V(\mathbf{z}^{(i)})\} - \min\{\partial_k V(\mathbf{z}^{(i)})\}, \quad k = 1, 2, \ldots, d. \end{aligned} \tag{15}$$

Our test calculations showed that scaling factors defined by Eq. (15) allow the recovery of the potential and derivatives at the data points while also being near-optimal for minimizing root-mean-square (rms) fitting errors.

The minimization conditions $\{\partial E/\partial a_j = 0\}$ lead to the modified normal equation:

$$\begin{aligned} &\left[\lambda_0^2(\mathbf{z})\mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{B} + \sum_{k=1}^{d} \lambda_k^2(\mathbf{z})\partial_k\mathbf{B}^{\mathrm{T}} W(\mathbf{z})\partial_k\mathbf{B}\right]\mathbf{a}(\mathbf{z}) \\ &\quad = \lambda_0^2(\mathbf{z})\mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\mathbf{f} + \sum_{k=1}^{d} \lambda_k^2(\mathbf{z})\partial_k\mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z})\partial_k\mathbf{f}, \end{aligned} \tag{16}$$

where matrices $\mathbf{B}$ and $\mathbf{W}(\mathbf{z})$ and vector $\mathbf{f}$ were defined earlier by Eqs. (4)–(6), matrix $\partial_k\mathbf{B}$ contains the kth partial derivative values of the basis functions at the data points:

$$\partial_k\mathbf{B} = \begin{pmatrix} \partial_k b_1(\mathbf{z}^{(1)}) & \partial_k b_2(\mathbf{z}^{(1)}) & \cdots & \partial_k b_m(\mathbf{z}^{(1)}) \\ \partial_k b_1(\mathbf{z}^{(2)}) & \partial_k b_2(\mathbf{z}^{(2)}) & \cdots & \partial_k b_m(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_k b_1(\mathbf{z}^{(n)}) & \partial_k b_2(\mathbf{z}^{(n)}) & \cdots & \partial_k b_m(\mathbf{z}^{(n)}) \end{pmatrix}, \tag{17}$$

and $\partial_k\mathbf{f}$ is the kth component of the gradient at the data points:

$$\partial_k\mathbf{f} = (\partial_k V(\mathbf{z}^{(1)}),\ \partial_k V(\mathbf{z}^{(2)}),\ \ldots,\ \partial_k V(\mathbf{z}^{(n)}))^{\mathrm{T}}. \tag{18}$$

Equation (16) is a simple generalization of Eq. (3) and can be solved in an identical way.

Higher order derivative data can be incorporated in the IMLS fits in a similar way. We will denote as IMLS-$D^{(\nu)}$ an IMLS fit that incorporates both energy and derivative data up to order ν. The work involved in the solution of the resulting normal equation is proportional to the total effective number of data values. For example, constructing the global IMLS-$D^{(2)}$ fit of n energy, nd gradient, and $nd(d+1)/2$ Hessian values for a d-dimensional PES will cost $(1 + 1.5d + 0.5d^2)$ times the cost of fitting energy values alone (the IMLS-$D^{(0)}$ fit). With cutoffs in the weight functions, however, the *effective local* value of n might very well decrease as higher order derivatives are added, perhaps amply compensating for the additional work required.

Others, e.g., Xie and Bowman [39], have incorporated gradients into global least-squares fits by generating additional "shadow" points in the neighborhood of the original (exact) data points via a finite difference approximation. The main advantage of generating shadow points rather than fitting the derivative values directly is that the derivatives of the basis functions are not required. However, choosing an appropriate finite differencing step size can be difficult to select for general application. Furthermore, if central differencing is involved, more than one shadow point is generated per derivative value, increasing the cost of constructing

Table 1 Summary of tested potentials

Potentials	$V_{MO}(z)$	$V_{HCO}(z)$
Definition	$V_{MO}(z) = D_e[1 - \exp(-\beta(z - z_0))]^2$	1-D slice of the HCO PES [a]
Constants	$\beta = 2a_0^{-1}$, $D_e = 100$ kcal/mol, $z_0 = 2a_0$	$\angle$ HCO = 170°; $R_{C-O} = 2.1a_0$
Variables	$1.653 \leq z/a_0 \leq 4.725$	$z = R_{C-H}$, $1.4 \leq z/a_0 \leq 4.472$
Dynamic	$0 \leq V_{MO}/(\text{kcal mol}^{-1}) \leq 100.3$	$2.3 \leq V_{HCO}/(\text{kcal mol}^{-1}) \leq 101.3$
Ranges	$-802 \leq V'_{MO}/(\text{kcal mol}^{-1}a_0^{-1}) \leq 100$	$-433 \leq V'_{HCO}/(\text{kcal mol}^{-1}a_0^{-1}) \leq 36$
Data sets	Nested grids of $n = 5, 9, 17, 33, 65, 129$, and 257 equidistant data points grid of $N = 1025$ equidistant evaluation points	

[a] Calculated at the B3LYP/6-311G(d,p) level of theory

and solving the IMLS normal equation relative to the direct approach.

4 Computational details

In this section we describe in order the PESs used to evaluate gradient incorporation, the basis set used in the IMLS, the weight function used in the IMLS and comparative modified Shepard calculations, and the data point selection schemes.

4.1 PES test cases

In order to evaluate various approaches to the incorporation of gradient data in the IMLS framework and better understand the properties of the interpolated potentials obtained by different methods, we have examined their performance for two representative 1-D potentials: the Morse oscillator (MO) potential and a reactive profile for constrained dissociation of the HCO radical to H and CO calculated at the B3LYP-DFT/6-311G(d,p) level of theory [40–43] with the Gaussian 03 program package [44]. Table 1 contains detailed parameters and characteristics of the tested potentials. The exact potentials and their first- and second-derivatives are shown in Fig. 1.

The Morse oscillator is chosen as a simple model of an anharmonic potential with a monotonic dissociation path, whereas the 1-D slice of the HCO PES represents a more complex dissociation path featuring an intermediate and a barrier. The latter was selected because there is a substantial barrier separating the HCO intermediate from H + CO products (see Fig. 1a). The HCO potential calculated at the B3LYP-DFT level of theory, while less accurate than some published HCO results [45,46], is representative of PES data generated directly by electronic structure calculations and thus an appropriate test case for our purpose. We have thoroughly and systematically tested the accuracy of different degree IMLS fits with and without gradient data, using root-mean-square (rms) errors in the fitted potential, gradient, and second derivatives calculated on a dense set of $N = 1,025$ evaluation points.

4.2 Basis set

We have used a standard polynomial basis as in our previous study [28] of 1-D applications of the IMLS method. We use the kd-IMLS notation when referring to the kth-degree IMLS fit, which by definition is constructed from the monomials of total degree less than or equal to k. In the 1-D case, the kd-IMLS basis contains $m = k + 1$ monomials: $b_1(z) = 1, b_2(z) = z, \ldots, b_m(z) = z^k$. Here we examined 1-D IMLS fits of up to fourth degree ($k = 1, 2, 3, 4$), taking the basis functions to be monomials shifted to the evaluation point, as suggested by Levin [25]. This choice of basis requires the $\mathbf{B}$ and $\partial_1\mathbf{B}$ matrices to be recalculated at every evaluation point, but it allows us to avoid large numbers in these matrices and simplifies Eq. (1) to $V_{\text{fit}} = a_1$, since $\mathbf{b}^{\text{T}} = (1, 0, \ldots, 0)$.

4.3 Weight function

In earlier work we examined several weight functions in conjunction with polynomial bases. Recently, we [31] and others [25] have analyzed in more detail cutoff strategies to make the fitting procedure more efficient by excluding remote data points that do not influence the accuracy of IMLS fits. These cutoff studies suggest a weight function of the form:

$$w(r_i) = s(r_i^2/R_{\text{cut}}^2)w_{\text{plain}}(r_i), \tag{19}$$

where $s(x)$ is a smooth damping function, such as:

$$s(x) = \begin{cases} \exp(-x/(1-x)) & \text{if } 0 \leq x < 1, \\ 0 & \text{if } x \geq 1, \end{cases} \tag{20}$$

and where $r_i = ||\mathbf{z} - \mathbf{z}^{(i)}||$ is the distance between the evaluation point $\mathbf{z}$ and the ith data point ($i = 1, 2, \ldots, n$), R_{cut} is a cutoff radius, and $x = (r_i/R_{\text{cut}})^2$. For a w_{plain} of infinite extent, the damping function s (x) makes w go smoothly to zero at $r_i \geq R_{\text{cut}}$ so that only those data points that are

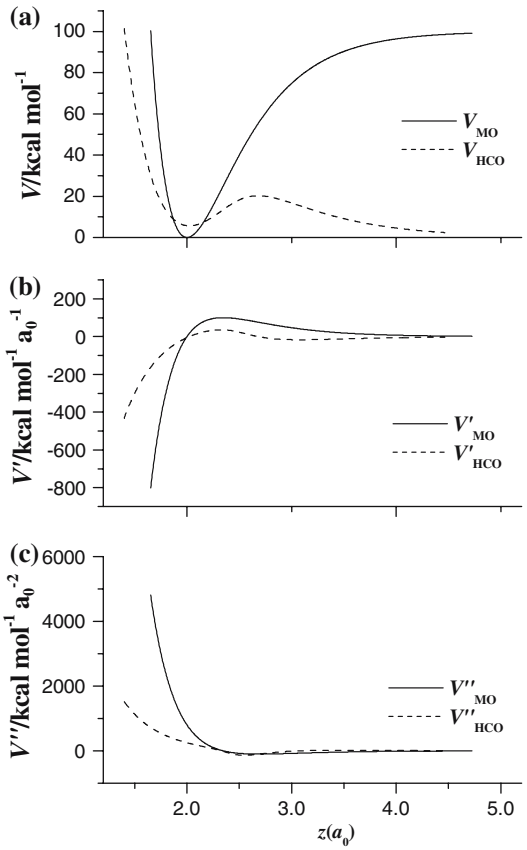

Fig. 1 **a** Values, **b** first and **c** second derivatives of the tested 1-D potentials

enclosed by R_{cut} are used to locally define the fit. To implement a cutoff strategy, one must have a method to determine R_{cut}. The simplest method, which is still shown to be effective [31], is to set R_{cut} to a fixed value, i.e., the Fixed Radius Cutoff (FRC) method, and we use it here. To apply Eq. (19) we must specify w_{plain} and R_{cut}.

The w_{plain} we use is adopted from the form suggested by McLain [19–22]:

$$w_{\text{plain}}(r_i) = \frac{\exp[-r_i^2/\zeta_i^2]}{(r_i^2/\zeta_i^2)^p + \varepsilon}, \tag{21}$$

where p is an integer, $p \geq 1$, ζ_i is a point-specific density adaptive parameter, and ε is a small positive number that prevents a numerical failure when $r_i \to 0$ and controls the dynamic range of weights. Lower p-values lead to smoother and globally more accurate fits; $p = 1$ was used here. The value of $\varepsilon = 1.\text{e}-7$ used here allows accurate reproduction of function values at the data points, while typically limiting the ratio of the largest to the smallest influential weight to less than 10^{16}, for numerical stability. The value of ζ_i controls how rapidly remote points are attenuated. Our numerical tests showed that accurate fits can be obtained by relating ζ_i's to $< h_i >$, the average data point spacing in the neighborhood of $z^{(i)}$, which can be defined such that $n_{\text{local}} < h_i >$ gives the length of the interval including $z^{(i)}$ and its n_{local} nearest neighbors. Here we used $n_{\text{local}} = 4$ to fit 1-D PESs with up to fourth degree polynomials. Then ζ_i is $< h_i >$ for IMLS and $< h_i >/2$ for IMLS-$D^{(1)}$. The lower value of ζ_i for IMLS-$D^{(1)}$ reflects the fact that incorporating derivatives allows a more attenuated local area to provide as accurate a value for the PES as a more extensive local area in IMLS without derivatives. With the ζ_i's set to unity, Eq. (21) is similar to other weight functions we have tried in previous IMLS studies. Those weight functions however had higher values of p to achieve sufficiently fast decay rates at long distances. Point-specific ζ_i's in Eq. (21) add more flexibility to the weight function because they allow the weight to adjust the long-range decay rate in direct response to the local density of points. The value of p can be selected to better control the short-range decay rate and the nature of singularity at $r_i = 0$. Smaller values of p give slower short-range decay rates, which in combination with density adaptive long-range decay controlled by the exponential factor in Eq. (21) give smoother and more accurate global fits. We use the minimal integer value allowed, $p = 1$. The fixed value of R_{cut} we use is:

$$R_{\text{cut}} = (n_{\text{local}} + 1)\max\{< h_i >\}. \tag{22}$$

This definition together with the procedure used to evaluate point-specific $< h_i >$'s guarantee that at least five energy data values are available to locally define up to 4d-IMLS fit at every point z: $\min\{z^{(i)}\} \leq z \leq \max\{z^{(i)}\}$. In general, if $n_{\text{local}} \geq k$ then the definitions of $< h_i >$'s and R_{cut} given here guarantee that at least $(k + 1)$ data points have non-vanishing weights and can be used to locally define the kd-IMLS fit using either equidistant or non-uniform data points.

4.4 Data point selection

Data point selection has been done by two different sampling methods: nested grids (GRID) and iterative automatic point selection (APS). The GRID method places data points on a series of nested grids with a mesh size progressively halved, starting with a coarse grid of $n_{\min} = 5$ and finishing with a fine grid of $n_{\max} = 257$ equally spaced data points. The same sets of data points are then used for all tested fitting methods. The GRID method allows systematic improvement in fitting accuracy by increasing the number of data points, although it is rather inefficient. Note that the rms errors are defined on a grid of 1,025 points, i.e., four times finer than the finest grid used in defining the ab initio points. Since IMLS fits have vanishingly small errors at the ab initio points included in

the fit, ever finer nested grids will by construction drive the rms error towards exceedingly small values.

An efficient data selection algorithm should place new data points in the regions where the fitting error is the largest. It is known that the IMLS fitting error is bounded in terms of the error of a local best polynomial interpolation [25], which in turn correlates with the variation between fits by different degree polynomials. While the fitting error is unknown a priori, a reasonable determination of where the fitting error is large can be made by finding where the difference between contending IMLS fits of different degree is the largest. We have used this approach to good effect in 1-D and higher applications of the standard IMLS method [28,30] and we use it here. Briefly, starting from $n_{\min} = 9$ equally spaced seed points, new data points were added iteratively (one at a time) from a dense global ensemble of evaluation points. For the kth-degree IMLS fit, each new data point was selected where the difference between the $(k-1)$th and kth-degree fits was the largest. The APS procedure was terminated when the maximum number of data points ($n_{\max} = 257$) was reached.

5 Test Calculations

In this section we examine the efficacy of including gradients along with values in an IMLS fit. The accuracy of the fit with respect to both its values and its derivatives will be examined. As discussed in Sect. 2, others have suggested approximations to the IMLS derivatives that are less expensive to evaluate. Since this is germane to our tests of the efficacy of incorporating gradients into IMLS fits, we will first examine these approximations. The rest of this section will display our evaluation of gradient incorporation.

5.1 IMLS derivative approximation

As mentioned in Sect. 2, the derivative of the IMLS fit [see Eq. (9)] has two terms, the first involving the derivatives of the basis functions and the second involving the derivatives of the IMLS coefficients. Others [5,25] have suggested that a useful approximation to the derivative of the IMLS fit is to set the second term to zero. There are two limits in which this is an excellent approximation: (1) when the vector of the derivatives of the IMLS coefficients $\partial\mathbf{a}$ is orthogonal to the vector of basis functions $\mathbf{b}$, and (2) when $\mathbf{a}$ is constant. The first limit is approached when n, the number of data points, is very large. One can show that vectors $\partial\mathbf{a}$ and $\mathbf{b}$ are orthogonal at the data points (see Appendix). As the density of data points increases with n, the $(\mathbf{b}^{\mathrm{T}}\partial\mathbf{a})$ term eventually becomes vanishingly small. The second limit is reached when the largest possible basis is used to fit limited data. For example, a minimum of $k+1$ data points is needed to define the kd-IMLS fit of a 1-D potential. In that limit the weight function becomes irrelevant and the least squares solution produces a globally constant $\mathbf{a}$. These two limits are illustrated in Fig. 2a for 2d- and 4d-IMLS fits for the MO case. In Fig. 2a the $(\mathbf{b}^{\mathrm{T}}\partial\mathbf{a})$ term measured as the rms difference between Eq. (9) and (11) evaluations of the IMLS derivative is plotted as a function of n for the GRID selection of data points. The rms representation of the $(\mathbf{b}^{\mathrm{T}}\partial\mathbf{a})$ term is essential zero for the minimum number of data points needed to determine the fits, $n_{\min} = 3$ and 5 for 2d- and 4d-IMLS, respectively. The largest difference between the one- and two-term IMLS derivative approximations is reached at $n = 9$ and it is decreasing toward zero as n grows large. Figure 2b shows the exact derivative and the computed derivatives by both Eqs. (9) and (11) over a partial range of z for $n = 17$. Eight of the 17 data points are found in this limited range of z and are indicated in the plot. The plots show that Eq. (11) typically produces an approximate derivative that is somewhat smoother than the exact IMLS derivative. Because of the absence of oscillations the approximate IMLS derivative is on average in better agreement with the true derivative than the exact IMLS derivative.

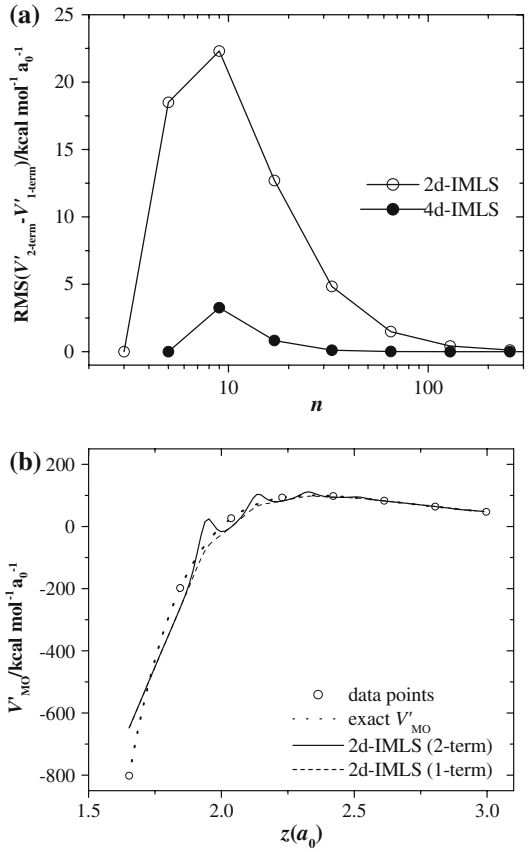

Fig. 2 V'_{MO} approximation from energy data on the grid. **a** RMS difference between the 1-term and 2-term derivative approximations by 2d- and 4d-IMLS as a function of n. **b** Exact V'_{MO} and its approximations from n = 17 energy data points by 2d-IMLS using the 1- and 2-term expressions, Eqs. (9) and (11); data points are indicated by *circles* on the exact V'_{MO} curve

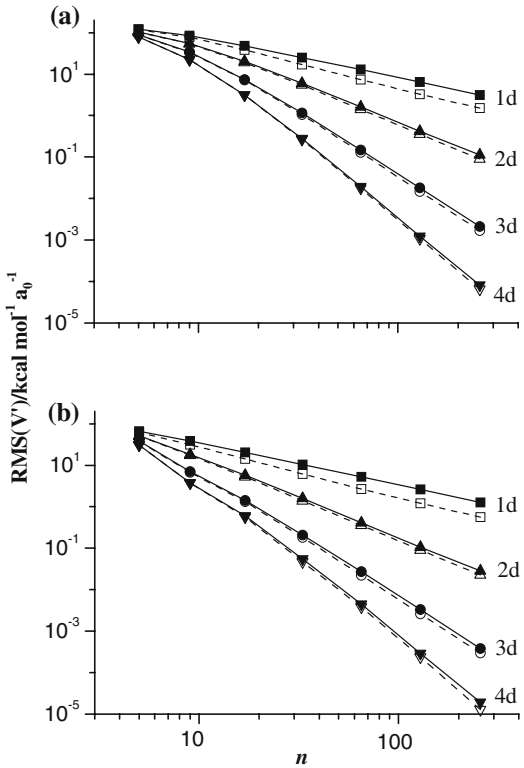

Fig. 3 **a** RMS errors for V'_{MO} and **b** V'_{HCO} approximated from energy values by different degree IMLS using the 1-term (*open symbols, dotted lines*) and 2-term (*filled symbols, solid lines*) expressions. Data points ($n = 5, 9, 17, 65, 129, 257$) were selected by the GRID method

Quite apart from the question of the error in the approximation relative to the exact IMLS derivative is the question of the error relative to the exact derivative on the true potential. The IMLS is itself an approximation to the true potential. An approximation such as Eq. (11) to the IMLS approximation can conceivably be in better agreement with the true derivative, as is seen in Fig. 2b. In Fig. 3 the rms error relative to the true derivative is shown as a function of n for IMLS derivatives calculated by the 1-and 2-term expressions, Eqs. (11) and (9), respectively. As in Fig. 2, the results are for nested grids and IMLS fits that do *not* incorporate gradients. The two panels of the figure represent our two test cases. As expected, the quality of the V' calculated by either method clearly improves with increasing degree of IMLS. However, the 1-term expression systematically gives a slightly better approximation whose superiority increases with increasing n or with decreasing IMLS degree.

One way to rationalize these results is as follows. The exact IMLS is "pinned" to each data point. The approximation of Eq. (11) if integrated from one data point to another would not be so closely pinned to the second data point. In other words, a fit constructed out of the integration of Eq. (11) fails to reproduce the accuracy of the exact IMLS at each data point but with the advantage of a "smoother" derivative as seen in Fig. 2b. By trading off accuracy at data points for derivative smoothness the integrated fit achieves somewhat greater derivative accuracy than the exact IMLS whose pinning at each data point is achieved by a more rapidly varying derivative between data points. The results in Fig. 3 show that the 1-term Eq. (11) provides an efficient and more accurate derivative approximation for both tested potentials. Although the results in the figure are for IMLS calculations that do not incorporate gradients, gradient incorporation does not qualitatively change the conclusions of the figure. In the following discussion the derivatives of IMLS-fitted PESs are calculated by Eq. (11) unless otherwise stated. Similarly, we have adopted a simple 1-term Eq. (12) to approximate the second derivatives of the potential. Their quality will be discussed in the next section.

Broader applications of the approximations of Eqs. (11) and (12) as well as the underlying reasons for their superior accuracy will require more study. While trajectory calculations seem an ideal application for Eq. (11), energy conservation, determination of product state distributions, and semiclassical eigenstate determinations all require both energies and derivatives. One potential problem that can be readily seen involves the geometry optimization on the fitted PES. Low-quality IMLS fits often have spurious minima and saddle points. For example, the exact 2d-IMLS derivative shown in Fig. 2b has three zeroes corresponding to two local minima separated by a small barrier, whereas the true derivative and the approximate one-term IMLS derivative [Eq. (11)] have only one zero that corresponds to the global minimum of V_{MO}. In this situation, geometry optimizations on the fitted PES using energy-only methods and conjugate-gradient methods will converge to different structures if the gradients are approximated by Eq. (11). Clearly, the implications of the inconsistency of derivatives via Eq. (11) and IMLS energies can be serious and require further systematic study. However, this inconsistency becomes negligibly small for more accurate IMLS fits using higher degree basis and denser sets of data points, as illustrated in Fig. 2a. Applications using IMLS fitted PESs should ensure sufficiently high quality of fits to avoid potential complications.

5.2 IMLS fits of energy and gradient data

The contrast in rms errors for the fitted potentials and their derivatives for IMLS with and without gradient incorporation is displayed in Figs. 4 and 5 for the two tested potentials: V_{MO} and V_{HCO}. The rms errors are plotted as a function of n in the GRID mode of data point selection. Each figure has four panels for representing the results of 1d-, 2d-, 3d-, and 4d-IMLS and IMLS-$D^{(1)}$ fits. The two figures show that, with the exception of the derivative for 1d-IMLS-$D^{(1)}$, incorporation of gradients lowers the rms error with respect

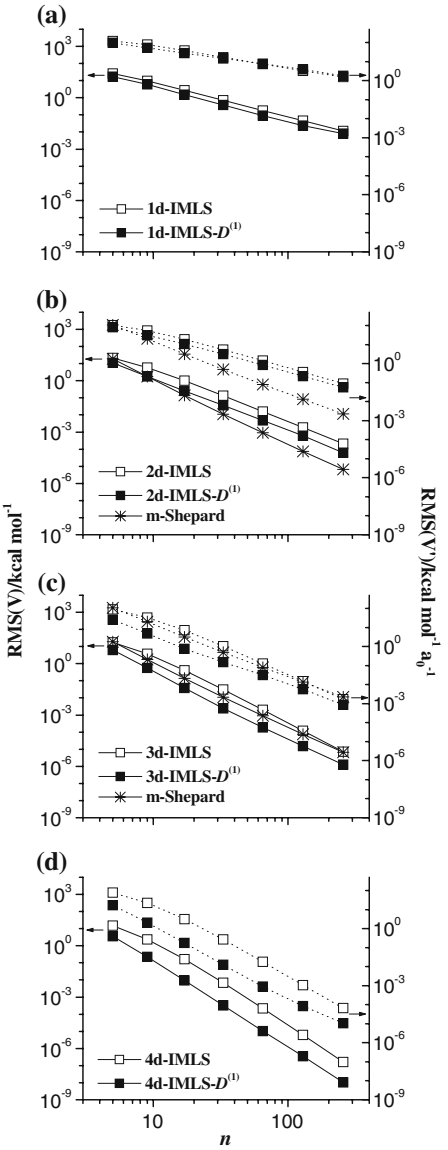

Fig. 4 RMS errors for V_{MO} (*solid lines*) and V'_{MO} (dotted lines) fitted by the **a** first, **b** second, **c** third, and **d** fourth degree IMLS (*open squares*), IMLS-$D^{(1)}$ (*filled squares*) and modified Shepard (*stars*) methods. Data points were selected by the GRID method

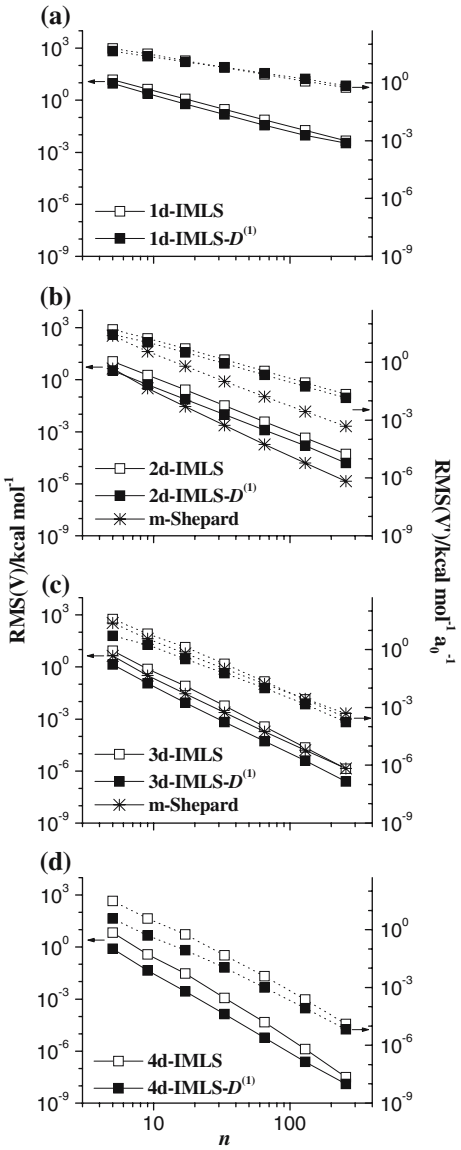

Fig. 5 RMS errors for V_{HCO} (*solid lines*) and V'_{HCO} (*dotted lines*) fitted by the **a** first, **b** second, **c** third, and **d** fourth degree IMLS (*open squares*), IMLS-$D^{(1)}$ (*filled squares*) and modified Shepard (*stars*) methods. Data points were selected by the GRID method

to the true PES. The results show the lowering of rms error is comparable for both value and derivative. With the exception of 1d-IMLS results where incorporating gradients causes minimal improvements or degradations in rms errors, all higher degree IMLS results show roughly comparable relative improvement in rms error. As expected, increasing the degree of IMLS systematically lowers the rms errors. In panels (b) and (c) of both Figs. 4 and 5, the modified Shepard fit is displayed as a function of n. To calculate this fit, each of the n data points supplies the computed value, gradient, and Hessian. As one can see in both Figs. 4 and 5 in panel (c), a 3d-IMLS-$D^{(1)}$ fit which involves no Hessians is superior in accuracy for both value and derivative to the modified Shepard fit. Comparison of panels (b) and (d) with panel (c) shows that 4d-IMLS and 4d-IMLS-$D^{(1)}$ fits are superior to modified Shepard while all 2d- and 1d-IMLS fits are not. In Fig. 6, the rms error of various 2d-, 3d-, and 4d-IMLS fits for the *second* derivative are compared as a function of n for the MO test case in panel (a) and the HCO test case in panel (b). The results are quite consistent with Figs. 4 and 5. Incorporating gradients improves the rms error by comparable relative amounts for each degree IMLS. Improving the degree systematically improves the error. By the 3d-IMLS the rms error becomes comparable or superior to modified Shepard.

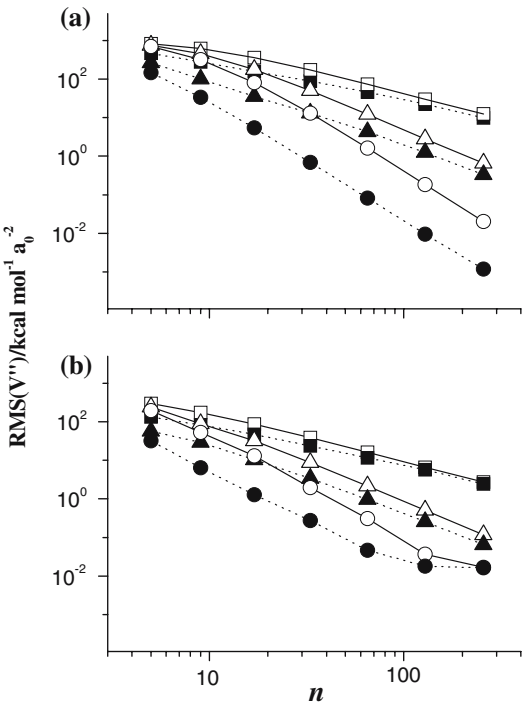

Fig. 6 RMS errors for **a** V''_{MO} and **b** V''_{HCO} approximated by different methods. Data points ($n = 5, 9, 17, 65, 129, 257$) were selected by the GRID method

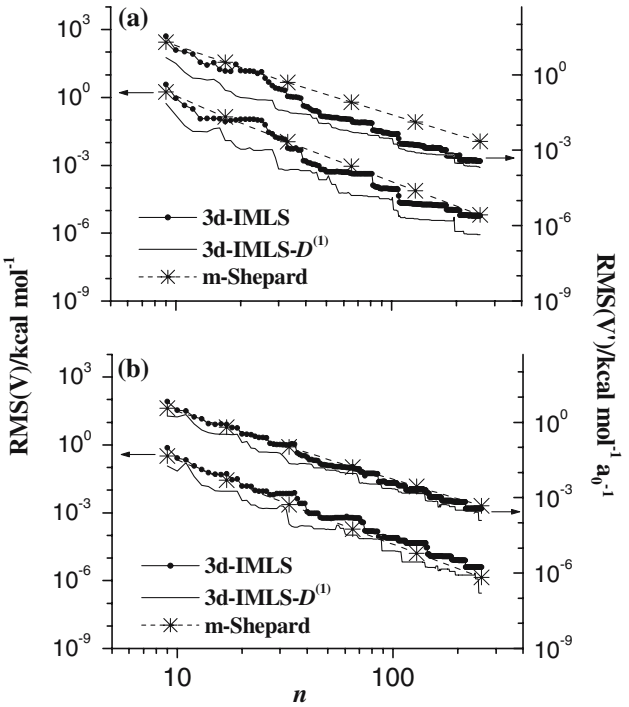

Fig. 7 RMS errors for **a** V_{MO} and V'_{MO}, **b** V_{HCO} and V'_{HCO} fits by 3d-IMLS and 3d-IMLS-$D^{(1)}$ as functions of the number of data points selected using the APS method. For comparison, comparable results of the modified Shepard method are shown as a function of the number of data points selected via the GRID method

Figure 7 shows the effect of gradient data incorporation at the 3d-IMLS level using the APS mode of data point selection. The rms errors in the fitted energies and derivatives are displayed as functions of the number of APS data points. Panels (a) and (b) are for the MO and HCO cases, respectively. In both panels the modified Shepard result is shown as a function of the number of data points selected via the GRID method. The modified Shepard results are identical to those in Figs. 4 and 5 and thus provide a convenient reference to gauge the impact of APS over GRID data point selection. The results in the figure, although less smooth, display the same qualitative trends as in Figs. 4 and 5. With respect to GRID selection, rms error is somewhat improved by APS. In particular, the modified Shepard results are equaled or surpassed by both 3d-IMLS-$D^{(1)}$ and 3d-IMLS. We observed similar effect of gradient data incorporation at the 4d-IMLS level. However, at lower levels (1d, 2d) of IMLS the convergence of APS is rather slow, especially for the derivative fits, which can be attributed to a poor quality of fitted derivatives at the low levels of IMLS.

A goal of this study is to estimate how many extra energy data points a derivative value is worth for improving the quality of fit. In Figs. 4 and 5 every symbol on a curve represents a factor of two increase in n. If an IMLS-$D^{(1)}$ rms error for some n_i in any panel of Figs. 4 and 5 were to equal the IMLS rms error for $2n_i$, then each gradient in the IMLS-$D^{(1)}$ fit does the "work" of one energy value in the IMLS fit. An examination of Figs. 4 and 5 shows that occasionally this limit is achieved, however, more often a derivative is worth two-thirds of an extra energy point. For many electronic structure methods, calculating all the gradients costs about as much as another energy calculation. For such methods, the results in Figs. 4 and 5 suggest calculating more energy values would be the most efficient approach to an accurate PES fit. However, this is a 1-D test case. These results are encouraging for multi-dimensional applications, even for 2-D. For those few electronic structure methods [36] where all the gradients cost only 10% that of computing a new energy, even in one-dimensional applications incorporating gradients is more efficient.

Because the results in Fig. 7 are not as smooth, it is difficult to apply the same analysis used immediately above for Figs. 4 and 5. To get a more realistic estimate we constructed the fits for V_{MO} and V_{HCO} using the same APS procedure as in Fig. 7 but with a termination when the rms difference between the contending IMLS fits reached selected values. We use two kinds of rms differences: one is the rms difference in energy and the second is the rms difference in derivative. The number of data points at termination with IMLS versus IMLS-$D^{(1)}$ tells us the effect of gradient incorporation in the APS data selection mode. These results are represented in Fig. 8 where panels (a) and (b) refer to the APS fits

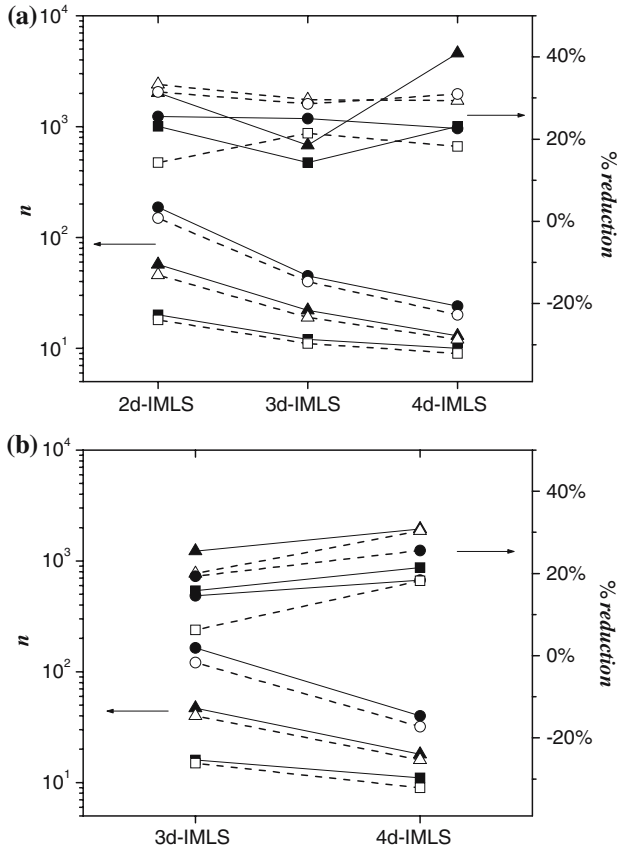

Fig. 8 As a function of the degree of the basis set, number of data points (selected via APS) required to achieve preset accuracy limits for an IMLS-$D^{(1)}$ fit and the percentage reduction in this number relative to that required for an IMLS fit with the same basis. *Solid symbols* and *lines* are for the MO test case, while *open symbols* and *dashed lines* are for the HCO test case. As described in the text, fitting accuracy for APS termination is measured by the rms difference between contending fits. In **a** the APS was terminated when the rms(V) difference dropped below 0.2 kcal/mol (*squares*); 0.02 kcal/mol (*triangles*), and 0.002 kcal/mol (*circles*). In **b** the APS was terminated when the rms(V') difference dropped below 2.0 kcal/(mol a_o) (*squares*); 0.2 kcal/(mol a_o) (*triangles*), and 0.02 kcal/(mol a_o) (*circles*)

terminated upon reaching the required accuracy limits in energy, and gradient, respectively. This figure shows that incorporating gradients improves the convergence of the APS energy and derivative fits by about 20–40%, which is comparable to the results obtained by the GRID method. Panel (b) shows only 3d and 4d IMLS results, because the convergence of 2d-IMLS APS derivative fits is slow due to the relatively poor quality of fitted derivatives at the 1d-IMLS-$D^{(1)}$ level. The result is large rms differences between the contending 1d- and 2d-IMLS fits.

6 Conclusions

We have combined ab initio gradient information with energy values to fit 1-D potentials using the interpolated moving least squares method. We have systematically compared results for various-order IMLS and the modified Shepard approaches. We have also applied an automatic point selection method to improve the accuracy of the IMLS fit. For the 1-D potentials using either APS or GRID data points the values of two or three gradients are on average comparable to an additional energy value in terms of the quality of the resulting IMLS fit. Since many electronic structure methods can calculated the full gradient vector in at least the time it takes to compute a scalar energy, these results suggest that for a PES of three dimensions or higher, an IMLS fit incorporating gradients would be cost effective. We expect for higher dimensional PESs gradient incorporation will typically be the approach of choice in IMLS applications. We have also discovered that for the 1-D cases we studied an approximation to the IMLS derivatives that ignores the variation of IMLS coefficients gives a better representation of the true derivative than the analytical IMLS derivatives. In future work we will examine gradient incorporation in multi-dimensional applications.

Acknowledgments This work was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Office of Science, US Department of Energy under Contract No. W-31-109-Eng-38 (Argonne) and Contract No. DE-FG02-01ER 15231 (MU). We have benefited from discussions concerning gradient incorporation with Dr. Richard Dawes (University of Missouri) and Professor Yin Guo (Oklahoma State University).

Appendix

In this appendix we prove that $\mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\partial_k \mathbf{a}(\mathbf{z}^{(i)}) = 0$ and derive formal error estimates for IMLS fitting methods. For convenience all results derived here will be only for IMLS methods but can be readily generalized to IMLS-$D^{(1)}$ methods. Several earlier studies [5,23,25] demonstrate that the standard IMLS normal equation, Eq. (3), has a unique solution and the resulting fit is interpolative [$V_{\mathrm{fit}}(\mathbf{z}^{(i)}) = V(\mathbf{z}^{(i)})$, where $\mathbf{z}^{(i)}$ is any data point] provided that the constant term is included in the basis (e.g., $b_1(\mathbf{z}) = 1$), matrix $\mathbf{B}$ [see Eq. (4)] has full rank, and the weight function has the following properties: $w(r) \geq 0$ and $w^{-1}(0) = 0$, where $r = ||\mathbf{z} - \mathbf{z}^{(i)}||$. The explicit form of this solution is given by Eq. (7). In deriving the error estimates for the IMLS method, it is useful to represent the interpolant in terms of cardinal functions $\{\psi_{0i}(\mathbf{z}), i = 1, 2, \ldots, n\}$:

$$V_{\mathrm{fit}}(\mathbf{z}) = \sum_{i=1}^{n} \psi_{0i}(\mathbf{z}) V(\mathbf{z}^{(i)}) = \boldsymbol{\psi}_0^{\mathrm{T}}(\mathbf{z})\, \mathbf{f}, \tag{A1}$$

where $\mathbf{f}$ is the column vector of the potential energy values at the data points [see Eq. (6)] and $\boldsymbol{\psi}_0(\mathbf{z}) = (\psi_{01}(\mathbf{z}), \psi_{02}(\mathbf{z}), \ldots, \psi_{0n}(\mathbf{z}))^{\mathrm{T}}$ is the vector of cardinal functions defined as:

$$\boldsymbol{\psi}_0^{\mathrm{T}}(\mathbf{z}) = \mathbf{b}^{\mathrm{T}}(\mathbf{z})\mathbf{S}_0^{-1}(\mathbf{z})\mathbf{B}^{\mathrm{T}}\mathbf{W}(\mathbf{z}), \tag{A2}$$

as can be seen from Eq. (7). The definitions of the vector $\mathbf{b}(\mathbf{z})$ and matrices $\mathbf{S}_0(\mathbf{z})$, $\mathbf{B}$, and $\mathbf{W}(\mathbf{z})$ were given in Sect. 2. IMLS cardinal functions minimize the quadratic form

$$Q(\mathbf{z}) = \sum_{i=1}^{n} \psi_{0i}^2(\mathbf{z})/w(||\mathbf{z}-\mathbf{z}^{(i)}||) = \boldsymbol{\psi}_0^{\mathrm{T}}(\mathbf{z})\, \mathbf{W}^{-1}(\mathbf{z})\boldsymbol{\psi}_0(\mathbf{z}), \tag{A3}$$

subject to the linear constraints:

$$\sum_{i=1}^{n} \psi_{0i}(\mathbf{z}) b_j(\mathbf{z}^{(i)}) = b_j(\mathbf{z}), \quad j = 1, 2, \ldots, m. \tag{A4}$$

Equivalently, $\mathbf{B}^{\mathrm{T}}\boldsymbol{\psi}_0(\mathbf{z}) = \mathbf{b}(\mathbf{z})$. The solution of this constrained minimization problem is exactly Eq. (A2), as can be verified using Lagrange multipliers [24,25]. Consequently, the IMLS interpolant can be constructed by solving either the IMLS normal equation, Eq. (3), or the constrained minimization problem formulated by Eqs. (A3) and (A4). An important property of IMLS fits that follows from Eq. (A4) is that any linear combination of basis functions is reproduced exactly by the fit.

The solution of Eqs. (A3), (A4) is very simple at the data points: $\psi_{0j}(\mathbf{z}^{(l)}) = \delta_{il}$ and $\boldsymbol{\psi}_0(\mathbf{z}^{(l)}) = \mathbf{e}_l$. Indeed, one can easily verify that: (1) $Q(\mathbf{z}^{(l)}) = \mathbf{e}_l^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{z}^{(l)})\mathbf{e}_l = w^{-1}(0) = 0$ is minimal and (2) $\sum_{i=1}^{n} \psi_{0i}(\mathbf{z}^{(l)}) b_j(\mathbf{z}^{(i)}) = \sum_{i=1}^{n} \delta_{il} b_j(\mathbf{z}^{(i)}) = b_j(\mathbf{z}^{(l)})$ for any basis function $b_j(\mathbf{z})$ and any data point $\mathbf{z}^{(l)}$.

Using the cardinality property of $\boldsymbol{\psi}_0$, it is easy to show that the vector $\partial_k\mathbf{a}$ is orthogonal to the vector $\mathbf{b}$ at the data points. Using Eqs. (10) and (A2) the product of $\partial_k\mathbf{a}$ and $\mathbf{b}^{\mathrm{T}}$ can be expressed in terms of cardinal functions:

$$\begin{aligned}\mathbf{b}^{\mathrm{T}}(\mathbf{z})\partial_k\mathbf{a}(\mathbf{z}) &= \mathbf{b}^{\mathrm{T}}(\mathbf{z})\mathbf{S}_0^{-1}(\mathbf{z})\mathbf{B}^{T}\partial_k\mathbf{W}(\mathbf{z})[\mathbf{f}-\mathbf{B}\mathbf{a}(\mathbf{z})] \\ &= \boldsymbol{\psi}_0^{\mathrm{T}}(\mathbf{z})(\partial_k\mathbf{W}^{-1}(\mathbf{z}))\mathbf{W}(\mathbf{z})[\mathbf{B}\mathbf{a}(\mathbf{z})-\mathbf{f}]\end{aligned} \tag{A5}$$

Then at the data points $\boldsymbol{\psi}_0(\mathbf{z}^{(i)}) = \mathbf{e}_i$ and

$$\begin{aligned}&\mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\partial_k\mathbf{a}(\mathbf{z}^{(i)}) \\ &= \partial_k w^{-1}(0) w(0)[\mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\mathbf{a}(\mathbf{z}^{(i)}) - V(\mathbf{z}^{(i)})] = 0,\end{aligned} \tag{A6}$$

because the inverse weight function has a minimum at $r = 0 (w^{-1}(0) = 0)$ so that $\partial_k w^{-1}(0) = 0$, and for any basis that includes $b_1(\mathbf{z}) = 1$ the minimization condition $\{\partial E_0/\partial a_1 = 0\}$ ensures that $w(0)[\mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\mathbf{a}(\mathbf{z}^{(i)}) - V(\mathbf{z}^{(i)})] = 0$.

This analysis can be extended to higher derivative expressions to show that $(\mathbf{b}^{\mathrm{T}}(\mathbf{z}^{(i)})\partial_l\partial_k\mathbf{a}(\mathbf{z}^{(i)})) = 0$. This extension is correct as long as higher derivatives of $w^{-1}(r)$ are zero at $r = 0$, such as $\partial_l\partial_k w^{-1}(0) = 0$ for every $k, l = 1, 2, \ldots, d$. Whether or not this is true depends on the specific form of the weight function.

The error in the IMLS approximation can be bounded in terms of the absolute values of the cardinal functions and the error of the best local approximation to $V(\mathbf{z})$ by a given basis set. Provided that $V(\mathbf{z})$ has bounded derivatives up to the $(k+1)$th order, 1-D IMLS fits using kth degree polynomial basis have the error that is bounded in terms of the local Lagrange interpolation error analysis:

$$|V(z) - V_{\mathrm{fit}}(z)| \le \left(1 + \sum_{i=1}^{n} |\psi_{0i}(z)|\right) \times \frac{\max\left(\left|\frac{\mathrm{d}^{k+1}V(z)}{\mathrm{d}z^{k+1}}\right|\right)\delta^{k+1}}{(k+1)!}. \tag{A7}$$

where δ is the range of data points used to locally define the fit. If the cutoff weight function [see Eq. (19)] is used, then δ is bound by the cutoff radius.

References

1. Ischtwan J, Collins MA (1994) J Chem Phys 100:8080
2. Franke R (1982) Math Comp 38:181
3. Franke R, Nielson G (1980) Int J Numer Methods Eng 15:1691
4. Farwig R (1987) Algorithms for approximation, Mason JC, Cox MG (eds) Clarendon, Oxford
5. Farwig R (1986) J Comput Appl Math 16:79
6. Farwig R (1986) Math Comput 46:577
7. Shepard D (1968) Proc. 23rd Nat Conf. ACM, New York, 517–524
8. Lancaster P, Salkauskas K (1986) Curve and surface fitting. An introduction, Academic, London, Chapter 10
9. Collins MA (2002) Theor Chem Acc 108:313
10. Jordan MJT, Thompson KC, Collin MA (1995) J Chem Phys 102:5647
11. Thompson KC, Collins MA (1997) J Chem Soc, Faraday Trans 93:871
12. Thompson KC, Jordan MJT, Collins MA (1998) J Chem Phys 108:8302
13. Thompson KC, Jordan MJT, Collins MA (1998) J Chem Phys 108:564
14. Betten RPA, Collin MA (1999) J Chem Phys 111:816
15. Crittenden DL, Thompson KC, Chebib M, Jordan MJT (2004) J Chem Phys 121:9844
16. Moyano GE, Collins MA (2004) J Chem Phys 121:9769
17. Ishida T, Schatz GC (1999) Chem Phys Lett 314:369
18. Ishida T, Schatz GC (2003) J Comput Chem 24:1077
19. McLain DH (1974) Comput J 17:318
20. McLain DH (1976) Comput J 19:178
21. McLain DH (1976) Comput J 19:384
22. Sabin MA (1976) Comput J 19:384
23. Lancaster P, Salkauskas K (1981) Math Comput 37:141
24. Bos LP, Salkauskas K (1989) J Approx Theory 59:267
25. Levin D (1998) Math Comput 67:1517
26. Wendland H (2001) IMA J Numer Anal 21:285
27. Maisuradze GG, Thompson DL (2003) J Phys Chem A 107:7118
28. Maisuradze GG, Thompson DL, Wagner AF, Minkoff M (2003) J Chem Phys 119:10002
29. Guo Y, Kawano A, Thompson DL, Wagner AF, Minkoff M (2004) J Chem Phys 121:5091
30. Maisuradze GG, Kawano A, Thompson DL, Wagner AF, Minkoff M (2004) J Chem Phys 121:10329
31. Kawano A, Tokmakov IV, Thompson DL, Wagner AF, Minkoff M (2006) J Chem Phys 124:054105
32. Kawano A, Guo Y, Thompson DL, Wagner AF, Minkoff M (2004) J Chem Phys 120:6414
33. Gaw JF, Yamaguchi Y, Schaefer HF (1984) J Chem Phys 81:6395

34. Gaw JF, Yamaguchi Y, Schaefer HF, Handy NC (1986) J Chem Phys 85:5132
35. Jørgensen P, Simons J (1986) Geometrical derivatives of energy surfaces and molecular properties. Reidel, Dordrecht
36. Lischka H, Shepard R, Pitzer RM, Shavitt I, Dallos M, Muller T, Szalay PG, Seth M, Yabushita GS, Kedziora GS, Zhang Z (2001) Phys Chem Chem Phys 3:664
37. Koizumi H, Schatz GC, Walch SP (1991) J Chem Phys 95:4130
38. Walch SP (1990) J Chem Phys 93:2384
39. Xie T, Bowman JM (2002) J Chem Phys 117:10487
40. Becke AD (1993) J Chem Phys 98:5648
41. Becke AD (1988) Phys Rev A 38:3098
42. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785
43. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) J Phys Chem 98:11623
44. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven JRT, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03, Revision B.04. Gaussian, Pittsburgh
45. Bowman JA, Bittman JS, Harding LB (1986) J Chem Phys 85:911
46. Werner HJ, Bauer C, Rosmus P, Keller HM, Stumpf M, Schinke R (1995) J Chem Phys 102:3593