Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes

Jie Cui and Roman V Krems

Department of Chemistry, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada

E-mail: rkrems@chem.ubc.ca

Received 16 March 2016, revised 11 September 2016
Accepted for publication 23 September 2016
Published 24 October 2016

Abstract
We explore the efficiency of a statistical learning technique based on Gaussian process (GP) regression as an efficient non-parametric method for constructing multi-dimensional potential energy surfaces (PESs) for polyatomic molecules. Using an example of the molecule N₄, we show that a realistic GP model of the six-dimensional PES can be constructed with only 240 potential energy points. We construct a series of the GP models and illustrate the accuracy of the resulting surfaces as a function of the number of ab initio points. We show that the GP model based on ∼1500 potential energy points achieves the same level of accuracy as the conventional regression fits based on 16 421 points. The GP model of the PES requires no fitting of ab initio data with analytical functions and can be readily extended to surfaces of higher dimensions.

Keywords: Kriging, potential energy surfaces, polyatomic molecules, Gaussian process regression

(Some figures may appear in colour only in the online journal)

Introduction

Accurate potential energy surfaces (PESs) describing the interactions of atoms in polyatomic molecules are required for calculations of the rovibrational energy levels and the dynamical properties of molecules. While classical dynamics calculations may use local values of the potential energy and its gradient, quantum dynamics calculations generally require the knowledge of the global PES. Obtaining the global PES for a polyatomic molecule requires quantum chemistry calculations of the potential energy at different molecular geometries and the construction of an analytical fit to interpolate the computed energies. The resulting PES must be continuous, smooth, differentiable and free of unphysical variations. As the complexity of the molecule increases, producing an analytical fit that satisfies these requirements becomes exceedingly difficult [1]. The difficulty arises from (i) the need to calculate the potential energy at a large number of coordinates; (ii) the complexity of the analytical functions and fitting procedures necessary for representing multi-dimensional surfaces.

Several methods have been developed recently to reduce the difficulty of fitting PESs for polyatomic molecules. For example, an n-mode representation was proposed to construct a PES as a series of intrinsic potentials accounting for a small number of normal modes [2–4]. An alternative approach—or one that can be used in combination with the n-node representation—is to take advantage of the symmetry of multi-dimensional PES under permutations of identical atoms [1, 5, 6]. The goal of these approaches is to fit a high-dimensional PES by fitting functions of lower dimensionality. However, both of these approaches are complex and, therefore, difficult to implement; are molecule-specific (i.e. the fitting procedure and the choice of analytical functions are different for different molecular species and different number of active degrees of freedom); and meet with the same challenges as the conventional fitting procedures when the dimensionality of the PES increases.

In order to construct accurate PESs for large polyatomic molecules or molecular complexes, it is desirable to develop alternative approaches that would be

(i) easy to implement;
This can be achieved by combining quantum chemistry calculations with machine-learning techniques developed for efficient interpolation of multi-dimensional spaces [7]. One of the simplest machine-learning methods is based on Shepard interpolation, giving a PES as a weighted average of the energy evaluated at different points with the weights decaying as simple functions of distance from the point of interest. For example, one can construct a multi-dimensional PES as a weighted sum of Taylor expansions around multiple points in the configuration space [8, 9]. While powerful and simple, Shepard interpolation is restricted by the form of the weight functions and may lead to results that are not monotonically converging. In addition, the approaches based on Taylor expansions require both the potential energies and the derivatives, as well as some prior knowledge of the PES energy landscape (e.g., the reaction path). Other machine-learning methods that can be used for obtaining high-dimensional PESs include an approach based on artificial neural networks [10] and Gaussian process (GP) regression [11]. Like Shepard interpolation, GP regression is a weighted sum of potential energy points at different coordinates but it is much more flexible, because the weights depend not only on the spatial but also on the statistical relationship between the points. It requires only the potential energy points on input and is expected to be much more efficient than Shepard interpolation. The application of artificial neural networks to fitting PESs has been explored in several studies [12–18], GP models have been proposed for constructing force-fields in [19–23] and for a variety of applications in molecular collision dynamics in [24, 25]. However, to the best of our knowledge, the accuracy and efficiency of GP regression for obtaining a global PES for a polyatomic molecule have not been systematically assessed before.

In the present work, we explore the efficiency of GP for interpolating multi-dimensional PESs by producing a series of GP models of the six-dimensional PES for N₂. We use the potential energy data computed by Paukku et al. [26] to train the GP models and explore the accuracy of the resulting surfaces as a function of the number of training points. This PES covers a wide range of energies between 0 and 1000 kcal mol⁻¹ and exhibits a complex energy landscape. We show that a realistic model representing accurately the low-energy part of the 6D PES can be obtained with only 240 quantum chemistry calculations. We illustrate that the GP model based on between 1200 and 1800 potential energy points achieves the same level of accuracy as the conventional regression fit obtained in previous work [26] based on 16 421 points.

Why GP for fitting PES?

Given the abundance of literature on the application of the artificial neural networks to fitting PESs [12–18], why to explore an alternative statistical learning technique? GP offer, at least, five major advantages for obtaining PESs:

(i) GP regression is a kernel-based statistical learning technique and, as such, is generally easier to implement than the artificial neural networks [19].

(ii) There is no need to fit any computed data with analytical functions [11, 27, 28]. A GP model is determined by correlations between potential energy points in a multi-dimensional configuration space and provides a statistical prediction for the value of the potential energy within the configuration space. The correlations do not need to be known exactly; rather, one uses the best estimates of the correlations parametrized by a simple analytical function.

(iii) With a proper parametrization of correlations between the energy points, a GP model is guaranteed to yield a smooth and differentiable surface that passes through the potential energy points used for training the model [29–32]. The choice of the correlation function controls the differentiability of the resulting surface.

(iv) The GP models scale favorably with the dimensionality of the problem. A rule of thumb is that a GP model should require on the order of \(10 \times q\) points for interpolating a surface with \(q\) dimensions [34]. This implies that a 6D PES can be obtained with only 60 quantum chemistry calculations. The purpose of this work is to test this rule of thumb in application to fitting the molecular PESs, typically characterized by smooth but wide variation of energy and the presence of multi-dimensional minima and barriers in the intramolecular coordinate space. Artificial neural networks are expected to require many more \(ab\ initio\) points [18].

(v) Techniques based on artificial neural networks often suffer from the problem of overfitting, i.e., including more potential energy points in the set used to train an artificial neural network may decrease the accuracy of the resulting PES fit [33]. By construction, GP regression does not have this problem so GP models are guaranteed to become more accurate when trained by more quantum chemistry calculations [34].

As will be clear from the discussion in the following section, the evaluation of the potential energy using a GP model is reduced to a product of a square matrix and a column vector with the dimension equal to the number of training points. Thus, the numerical efficiency of the PES evaluation based on a GP model scales as \(\sim(p \times q)^2\) with the dimensionality \(q\) of the PES and the effective number \(p\) of \(ab\ initio\) points per dimension. This may become a bottleneck when \(p \times q \approx 10^3–10^4\). One should, therefore, expect the GP regression to be a method of preference for dimensions \(q \lesssim 100\) and the artificial neural network approach—for problems with large \(q \gtrsim 100\).
Fitting PES with a GP

We will denote a GP by \( F(\cdot) \). A GP is a family of normally distributed random multi-dimensional functions characterized by a mean function \( m(\cdot) \) and a covariance function \( K(\cdot, \cdot) \). For a GP with constant variance \( s^2 \), the covariance function is \( K_R, s^2 = K(\cdot, \cdot) \), where \( R(\cdot, \cdot) \) is a correlation function. A point in a multi-dimensional space of interest is specified by a vector \( x \). A realization of a GP at a given \( x \) is the value of a random function drawn from the normal distribution and evaluated at \( x \). Multiple outputs of a GP at the same \( x \) produce a Gaussian distribution of values \( F(x) \).

The application of the GP models to molecular dynamics problems has been described in our previous work \([24, 25]\). Here, we assume that the internal coordinates of a polyatomic molecule are given by a \( q \)-dimensional vector \( x\). Our goal is to construct the global PES, given \( n \) potential energy values at vectors \( x, \ldots, x_n \). We refer to these \textit{ab initio} points as the training points. We model the collection of \( n \) points in the \( q \)-dimensional space by a GP and assume that each of these potential energies \( V(x_i) \) computed by a quantum chemistry method is a realization of a GP at \( x_i \).

The multiple outputs of a GP at the given \( n \) points \( Y^n = (F(x_1), F(x_2), \ldots, F(x_n))^T \) follow a multivariate normal distribution

\[
Y^n \sim \text{MVN}(\beta, \sigma^2 A),
\]

where \( \beta \) is the mean vector and \( A \) is a \( n \times n \) matrix defined as

\[
A = \begin{pmatrix}
1 & R(x_1, x_2) & \cdots & R(x_1, x_n) \\
R(x_2, x_1) & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
R(x_n, x_1) & \cdots & 1
\end{pmatrix}.
\]

In equation (1) as well as equation (8) the \( \sim \) sign means ‘distributed as’. For simplicity, and without loss of generality, we assume that \( \beta = \beta I \), where \( I \) is the identity vector of dimension \( n \) and \( \beta \) is an unknown scalar parameter.

We describe the correlation function \( R(\cdot, \cdot) \) by the following expression \([29–32]\):

\[
R(x, x') = \prod_{i=1}^{q} \left( 1 + \frac{\sqrt{5}|x_i - x'_i|}{\omega_i} + \frac{5|x_i - x'_i|^2}{3\omega_i^2} \right) \times \exp \left( -\frac{\sqrt{5}|x_i - x'_i|}{\omega_i} \right),
\]

where \( \omega_i \) are the unknown parameters representing the characteristic length scales of the correlations. Equation (3) is a special case of the Matérn correlation function \([29–32]\) defined as

\[
R(x, x') = \prod_{i=1}^{q} \frac{1}{\Gamma(\nu)2^{\nu-1}}d_i^{\nu-1}K_{\nu}(d_i),
\]

where \( d_i = \sqrt{2\nu|x_i - x'_i|}/\omega_i \) and \( K_{\nu}(\cdot) \) is the modified Bessel function of order \( \nu \). Equation (3) represents the Matérn correlation function with \( \nu = 5/2 \). The choice of the mathematical form of the correlation function is not unique. The

\[\text{Figure 1.} \text{ Intermolecular geometries used for the computation of the PES for N}_4 \text{ in [26].}\]
The mathematical form of the correlation function determines the efficiency of the GP model (i.e. the number of training points required to obtain a GP model of a given accuracy) as well as the differentiability of the resulting surface. For example, if the correlation function is chosen to be a Gaussian \([29–32]\), the GP is differentiable to any order. With the Matérn correlation function \((4)\), the process is differentiable to order \(k < \nu\). Thus, when the parameter \(\nu = 5/2\), the GP is twice differentiable. For the application of the present work, we have tested several different correlation functions and found the Matérn function to yield the most accurate results for a fixed number of training points. For any given molecule, one should always try different correlation functions to find the most optimal one. For example, we have found that for other molecular systems, a Gaussian correlation function yields more accurate results for a fixed number of training points.

To find the estimates of the parameters \(\omega = (\omega_1, \omega_2, \cdots, \omega_p)\) giving the most likely correlation
function for the potential energy points available, known in statistics as the maximum likelihood estimates (MLE), we maximize numerically the log-likelihood function

$$\log L(\omega|Y^n) = -\frac{1}{2} [n\log\hat{\sigma}^2 + \log(\det(A)) + n], \quad (5)$$

where

$$\hat{\sigma}^2(\omega) = \frac{1}{n}(Y^n - \beta)^T A^{-1}(Y^n - \beta), \quad (6)$$

$$\hat{\beta}(\omega) = (T^TA^{-1}T)^{-1}T^TA^{-1}Y^n, \quad (7)$$

and the hat over the symbol denotes the maximum likelihood estimators. We find the maximum of equation (5) by iteratively computing the inverse of A and its determinant and using the Broyden–Fletcher–Goldfarb–Shanno optimization algorithm with a maximum of 100 iterations.

The goal is to make a prediction of the potential energy value at an arbitrary position $x = x_0$. The values $Y_0 = F(x_0)$ obtained by multiple realizations of the GP at $x_0$ and the multiple outputs of the GP at training sites $Y^n = (F(x_1), F(x_2), \cdots, F(x_n))$ are jointly distributed as

$$\begin{pmatrix} Y_0 \\ Y^n \end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} 1 \\ I \end{pmatrix} \beta, \sigma^2 \begin{pmatrix} A_0 & A \ \\ A & A \end{pmatrix} \right), \quad (8)$$

Figure 3. Same as in figure 2 but for the PES for $N_4$ represented by the GP model trained with 1200 potential energy points.
where $A_0 = (R(x_0, x_1), R(x_0, x_2), \ldots, R(x_0, x_n)^T$ is a column vector specified by the correlation function $R(\omega)$ with the MLE of $\omega$. This means that the conditional distribution of values $y = F(x_0)$ given the values $Y^n$ is a normal distribution with the conditional mean

$$\mu(x_0) = \beta + A_0^T \Lambda^{-1}(Y^n - \beta)$$

and the conditional variance

$$\sigma^2(x_0) = \sigma^2(1 - A_0^T \Lambda^{-1}A_0).$$

Once the MLE of the correlation matrix $\Lambda$ is determined by the maximization of the log-likelihood function, we use equation (9) as a prediction of the value of the potential energy at $x_0$ and equation (10) as the uncertainty of the prediction.

### Results

The PES for $N_4$ considered in this work was calculated in [26] for multiple $N_2-N_2$ and $N_3-N$ geometries shown in figure 1. Combined with the radial dependence of the potential energy for each of these geometries, these data provide a set of potential energy points in six-dimensions, covering a wide range of energies between 0 and 1000 kcal mol$^{-1}$. We use subsets of these $ab\ initio$ points to construct the GP models of the PES. In principle, the potential energy points in vector $Y^n$ used for training the GP model can be chosen at random configurations of the molecule $x_i$. However, given a fixed number of training points, the accuracy and the stability of the GP model can be improved if the points are selected to cover evenly the configuration space of interest. This can be achieved, for example, with the Latin hypercube sampling (LHS) method [35] known to cover a multi-dimensional space efficiently. In order to improve the efficiency of the GP models, we use the following sampling technique for the PES of $N_4$. Paukku et al. [26] computed 15363 points for 9 combinations of internal angles (9 sets of geometries) for $N_4$ and 1056 points for 77 combinations of internuclear distances for $N_3$. We first treat each combination of three input variables as a single variable to get a reduced 4D PES. We then generate an $n \times 4$ LHS matrix with values uniformly distributed on $[0, 1]$, and transform elements in each row to sample quantiles corresponding to the empirical distribution of input variables in the reduced 4D PES. If the sampled configuration is not contained in the data set of Paukku et al. [26], we use the nearest available configuration. It would be better to use the LHS in six-dimensions. However, that would lead to a lot of configuration points not contained in the data of Paukku et al. [26]. The sampling technique used here requires a much smaller number of adjustments to the sampled configurations.

Following [26], we illustrate the quality of the PES model by figures 2 and 3 showing the potential energy curves for the dissociation $N_4 \rightarrow N_3 + N$ at different values of $d$ of the distance between the centers of mass of the two nitrogen molecules. The different panels of figures 2 and 3 correspond to different geometries of the $N_4$ complex, covering the $A$-shaped, $T$-shaped, $H$-shaped and $X$-shaped sets illustrated in figure 1. The symbols in figures 2 and 3 represent the potential energy data and the curves are the GP models. Note the difference between figures 2 and 3 illustrating the improvement of accuracy of the GP model of the PES as the number of training points is increased.

Figure 2 is obtained with only 240 potential energy points. Given the small number of $ab\ initio$ points, the agreement between the symbols and the GP model curves is remarkable. Figure 2 illustrates that 240 potential energy points produce a GP model representing accurately the low-energy part ($\leq 200$ kcal mol$^{-1}$) of the global 6D PES and giving a qualitatively correct representation of the PES at high energies ($\geq 200$ kcal mol$^{-1}$). Figure 3 obtained with 1200 $ab\ initio$ points illustrates a similar level of accuracy as the fit in [26]. See table 1 for the quantitative information on the accuracy of the PES models. We note that it takes very little computing time to construct the GP models shown in figures 2 and 3. Training the models involves the iterative inversion of the correlation matrix $\Lambda$ that has the dimension $n \times n$, where $n$ is the number of training points, i.e. 240 and 1200 in the examples here. Since we use less than 100 iterations to maximize the log-likelihood function, inverting the matrices presents no computational difficulty. On the other hand, evaluating the PES using equation (9) involves a

| Number of points | MUE (low E) | RMSE (low E) | MUE (all E) | RMSE (all E) |
|------------------|-------------|--------------|-------------|--------------|
| PES of [26]      | 16421       | 1.3          | 1.8         | 5.0          | 14.3         |
| GP model         | 240         | 4.43 ± 1.37  | 6.39 ± 1.77 | 13.97 ± 2.02 | 38.00 ± 7.92 |
|                  | 480         | 2.61 ± 0.99  | 4.19 ± 1.56 | 9.68 ± 1.05  | 28.67 ± 3.92 |
|                  | 720         | 1.72 ± 0.52  | 2.96 ± 0.78 | 7.72 ± 0.78  | 24.02 ± 3.15 |
|                  | 960         | 1.39 ± 0.34  | 2.58 ± 0.67 | 6.57 ± 0.77  | 21.57 ± 2.78 |
|                  | 1200        | 1.23 ± 0.28  | 2.26 ± 0.49 | 5.76 ± 0.59  | 19.46 ± 2.75 |
|                  | 1800        | 0.96 ± 0.23  | 1.81 ± 0.46 | 4.45 ± 0.57  | 15.87 ± 2.72 |
|                  | 2400        | 0.74 ± 0.19  | 1.50 ± 0.37 | 3.83 ± 0.53  | 14.33 ± 2.38 |
matrix—vector multiplication, which scales as $O(n^2)$, making the evaluation of the PES trained by 1200 points 25 times more time consuming than the evaluation of the PES trained by 240 points. This may be a computational bottleneck in an application requiring a large number of PES evaluations and needs to be taken into consideration, when applying GP regression.

The comparison of figures 2 and 3 shows that increasing the number of potential energy points improves the quality of the model both at low energies and in the high energy region of the surface. In order to quantify the accuracy of the GP models of the PES and compare the result with those in [26], we calculate the mean unsigned error (MUE) and the root mean square error (RMSE) for a series of GP models obtained with different numbers of potential energy points. The errors are defined as

$$\text{MUE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|, \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}, \quad (12)$$

where $\hat{y}_i$ is the value of energy obtained from the GP model and $y_i$ is the corresponding value of energy from the data of Paukku et al [26]. To ensure that the accuracy of the GP models shown in figures 2 and 3 is not accidental, we construct up to 100 different GP models for each number of ab initio points summarized in table 1 and compute the average errors with the corresponding standard deviations of

*Figure 4.* Three-dimensional cross sections of the 6D PES for $N_4$ produced by the Gaussian Process model trained with 1800 points. Upper left panel: dependence of the PES on the angles $\theta_A$ and $\theta_B$ with $d = 1.4$ Å, $r_A = 1.098$ Å, $r_B = 1.5$ Å. Upper right panel: dependence of the PES on the angles $\theta_A$ and $\phi$ with $d = 1.4$ Å, $r_A = 1.098$ Å, $r_B = 1.5$ Å, $\theta_B = 60$ degrees. Lower panel: Bent-T-shaped configuration of figure 1 with $\phi = 0$, $\theta_A = 0$, $\theta_B = 60$ degrees and $r_A = 1.098$ Å. The angles and the geometries are defined in figure 1.
the errors. The number of the GP models was chosen to ensure convergence of both the mean values and the standard deviations reported in table 1.

Table 1 illustrates that the GP model trained by 1800 potential energy points produces the PES with the similar accuracy as the fitting procedure of [26] based on polynomial regressions and 16-421 potential energy points. Table 1 also illustrates the monotonous improvement of both the model accuracy and the model stability as the number of potential energy points increases. The accuracy of the GP model increases by almost 40% as the number of potential energy points increases from 240 to 480, and another 50% as the number of points increases from 480 to 1800. The increase of the stability is evident from the decrease of the standard deviations. The data in table 1 illustrate that the low-energy range between 0 and 100 kcal mol\(^{-1}\) can be represented with the RMSE \(< 2\) kcal mol\(^{-1}\) by a GP model with only 720 potential energy points.

In order to illustrate that the GP models described here interpolate smoothly the \textit{ab initio} points in all six-dimensions, we present in figure 4 three-dimensional cuts showing the dependence of the potential energy on the radial coordinates as well as the angles defined in figure 1. The PES shown in figure 4 is constructed with 1800 \textit{ab initio} points. The plots in figure 4 illustrate that the PES approaches proper asymptotic limits and is free of cusps or discontinuities, which are often inevitable if the PES is constructed by joining analytical functions representing different part of the surface. The structure at short range in the lower panel of figure 4 can be more clearly seen in figures 2 and 3.

Conclusion

We have constructed a series of GP models of the global 6D potential energy surface for the molecule N\(_4\) and showed that an accurate surface can be obtained with only \sim\1500 potential energy points. The number of potential energy points required for an accurate GP model of the PES can be decreased by selecting the points to follow a 6D LHS [35] instead of a quasi-LHS used here and by using regression functions to approximate the unconditional mean function \(\mu(\cdot)\). For example, the mean of the GP can be modeled as

\[
\mu(x) = \sum_{j=1}^{s} h_j(x)\beta_j = h(x)^T\beta, \tag{13}
\]

where \(h = (h_1(x), ..., h_s(x))^T\) is a vector of \(s\) regression functions [27, 28] chosen to mimic the dependence of the PES on the corresponding variables. Since the variation of the PES with the individual bond lengths is often nearly quadratic at low energies and follows simple power laws in the limit of bond dissociations, it should often be possible to find suitable regression functions \(h_j\) to optimize the GP model training. We note that the final results do not depend on the specific form of the functions \(h_j(x)\), as is clear from the calculations in the present article based on the simple choice of \(h_1 = 1\) and \(h_{j>1} = 0\). The efficiency of GP models can also be increased when combined with the sampling schemes based on classical trajectory calculations, as proposed in [8]. Since GP regression produces a qualitatively correct PES with already a small number of randomly selected points (240 for the 6D example considered here), one can envision, inspired by [8], an iterative procedure based on the following steps: (i) construct a crude PES with a small number of points; (ii) run a number of classical trajectories on the crude PES to determine the parts of the configuration space most frequently visited; (iii) add more points to those parts of the configuration space; (iv) re-construct the GP model; (v) repeat steps (ii) through (v).

We emphasize that the method of obtaining PES discussed here is completely general and can be applied to any polyatomic system. There are no assumptions made about the type or character of the molecular system. As a supervised learning technique, GP regression is guaranteed to become more accurate when trained by more \textit{ab initio} points [34]. This is illustrated by table 1. In this work, we performed calculations for the specific molecule N\(_4\). We have chosen this system because the PES for N\(_4\) spans a very wide range of energies (between 0 and 1000 kcal mol\(^{-1}\)) and exhibits a complex energy landscape. It is usually more difficult to construct accurate fits of PES covering a wider range of energies. To illustrate the generality of the approach described here, we constructed the GP models of the PES for two other molecular systems: H\(_3\) and HFCO [36]. For H\(_3\) we obtained a GP model representing the three-dimensional PES at energies \(\pm 46\) kcal/mol with RMSE \(< 0.1\) kcal mol\(^{-1}\) based on 210 potential energy points. For HFCO, we obtained a GP model representing the global six-dimensional PES spanning the energy range between 0 and \sim\57 kcal mol\(^{-1}\) with MUE 0.70 \pm 0.07 and RMSE 1.19 \pm 0.11 kcal mol\(^{-1}\) based on 300 \textit{ab initio} points. Increasing the number of training \textit{ab initio} points to 800 decreases the RMSE of the global 6D PES for HFCO to \approx 0.48 \pm 0.06 kcal mol\(^{-1}\).

Our results have two important implications. First, a GP model, even with the simple choice of \(h_1 = 1\) and \(h_{j>1} = 0\), requires much fewer potential energy calculations for the construction of an accurate global potential surface than conventional fitting techniques based on polynomial regressions. Second, the extension of the GP model discussed here to surfaces of higher dimensions is straightforward. Adding another dimension simply requires an extension of the expression for the correlation function to include another term in the product (3). Adding another dimension also requires more training points in the input vector \(Y^0\). However, it is known that the number of additional training points required per added dimension decreases with the number of dimensions [34]. The numerical procedure of training the GP model is limited by inversion of the matrix \(A\). The current limit on the number of dimensions amenable to GP regression is about 100 [37] and there is currently active research on extending the applications of GP models to problems of higher dimensionality [38]. It is thus foreseeable that the GP model discussed here can be applied for constructing the PES for molecules with up to 100 internal degrees of freedom. Such a PES model would require an estimated 10 000–30 000 potential energy points, which is currently well within reach.
of state-of-the-art quantum chemistry calculations. Please note that, however encouraging, this estimate is based only on what is know about the general improvement of GP regression with increasing dimensionality and the number of points for an accurate model of the 6D molecular system demonstrated here. For PES with sharp features, such as conical intersections, the number of \textit{ab initio} calculations required may be significantly larger. We emphasize, however, that the same computer code can be used to obtain PES with any dimensionality so there is no extra programming effort necessary to extend the present work to systems of many more dimensions.

Acknowledgments

We thank Dong Hui Zhang for providing the PES for H$_3$ and Alex Brown for providing the PES for HFCO. This work is supported by NSERC of Canada.

References

[1] Braams B J and Bowman J M 2009 Int. Rev. Phys. Chem. 28 577
[2] Carter S, Culik S J and Bowman J M 1997 J. Chem. Phys. 107 10458
[3] Yagi K, Oyanagi C, Taketsugu T and Hirao K J 2003 J. Chem. Phys. 118 1653
[4] Rauhut G, Barone V and Schwerdtfeger P 2006 J. Chem. Phys. 125 054308
[5] Wang Y, Braams B J, Bowman J M, Carter S and Tew D P 2008 J. Chem. Phys. 128 224314
[6] Conte R, Qu C and Bowman J M 2015 J. Chem. Theory Comput. 11 1631
[7] Rupp M 2015 J. Int. Quantum Chem. 115 1058
[8] Ischtwan J and Collins M A 1994 J. Chem. Phys. 100 8080
[9] Collins M A 2002 Theor. Chem. Acc. 108 313
[10] Hornik K, Stinchcombe M and White H 1989 Neural Netw. 2 359
[11] Sacks J, Schiller S B and Welch W J 1989 Technometrics 31 41
[12] Raff L M, Malshe M, Hagan M, Doughan D I, Rockley M G and Komanduri R J 2005 J. Chem. Phys. 122 084104
[13] Manzhos S, Wang X, Dawes R and Carrington T 2006 J. Phys. Chem. A 110 5295
[14] Handle C M and Popelier P L A 2010 J. Phys. Chem. A 114 3371
[15] Behler J and Parrinello M 2007 Phys. Rev. Lett. 98 146401
[16] Darley M G, Handle C M and Popelier P L A 2008 J. Chem. Theory Comput. 4 1435
[17] Behler J 2011 Phys. Chem. Chem. Phys. 13 17930
[18] Chen J, Xu X, Xu X and Zhang D H 2013 J. Chem. Phys. 138 221104
[19] Handle C M, Hawe G L, Kellab D B and Popelier P L A 2009 Phys. Chem. Chem. Phys. 11 6365
[20] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Phys. Rev. Lett. 104 136403
[21] Bartok A P and Csányi G 2015 Int. J. Quantum Chem. 115 1051
[22] Caccin M, Li Z, Kermode J R and Vita A De 2015 Int. J. Quantum Chem. 115 1129
[23] Popelier P L A 2015 Int. J. Quantum Chem. 115 1005
[24] Cui J and Krems R V 2015 Phys. Rev. Lett. 115 073202
[25] Cui J, Li Z and Krems R V 2015 J. Chem. Phys. 143 154101
[26] Paukku Y, Yang K R, Varga Z and Truhlar D G 2013 J. Chem. Phys. 139 044309
[27] Santner T J, Williams B J and Notz W I 2003 The Design and Analysis of Computer Experiments (New York: Springer Science & Business Media)
[28] Rasmussen C E and Williams C K I 2006 Gaussian Process for Machine Learning (Cambridge, MA: MIT Press)
[29] Mitchell T, Morris M and Ylvisaker D 1990 Stoch. Process. Appl. 35 109
[30] Cressie N A C 1993 Statistics for Spatial Data (New York: Wiley-Interscience)
[31] Stein M L 1999 Interpolation of Spatial Data: Some Theory for Kriging (New York: Springer Science Business Media)
[32] Abt M 1999 Scand. J. Stat. 26 563
[33] Zhang D H and Guo H 2016 Annu. Rev. Phys. Chem. 67 135
[34] Loepky J L, Sacks J and Welch W J 2009 Technometrics 51 366
[35] Stein M 1987 Technometrics 29 143
[36] Pradhan E and Brown A 2016 J. Mol. Spectrosc. in press
[37] Dubourg V, Sudret B and Deheeger F 2013 Probabilist. Eng. Mech. 33 47
[38] Linkletter C, Bingham D, Hengartner N, Higdon D and Ye K Q 2006 Technometrics 48 478