ORIGINAL PAPER

A Monte Carlo error estimator for the expansion of rigid-rotor potential energy surfaces

Claire Rist · Alexandre Faure

Abstract   The computation of rotational energy transfer in nonreactive molecular collisions requires expanding the orientation dependence of the interaction potential over an appropriate complete set of orthonormal functions. We show that the use of random grids for the sampling of the angular geometries combined with the Monte Carlo theorem allows to estimate the mean accuracy on each expansion term determined by a least squares fit. The interest of our approach is illustrated by an application to the H_2O–H_2 system, of great astrophysical interest.

Keywords   Rotational energy transfer · Potential energy surface · Angular expansion · Monte Carlo theorem

1 Introduction

The most accurate method for calculating rotational energy transfer in (nonreactive) molecular collisions is the quantum close coupling approach. This method consists in solving the time independent Schrödinger equation by expanding the total wave function in a basis set of rotation functions of the colliding partners and partial wave expansion for the collision coordinate. In the resulting coupled second-order differential equations, the coupling comes from the orientation dependence of the interaction potential, or potential energy surface (PES). The computation of the necessary matrix

C. Rist · A. Faure (✉)
UJF-Grenoble 1/CNRS, Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) UMR 5274, 38041 Grenoble, France
e-mail: afaure@obs.ujf-grenoble.fr

elements of the potential over the expansion basis functions is usually performed by expanding the orientation dependence of the PES over a complete set of appropriate analytical orthonormal functions, see e.g. [1].

High accuracy PES are generally obtained using *ab initio* quantum chemistry methods such as the coupled cluster theory, see e.g. [2,3] for the $H_2O–H_2$ system. In order to reproduce all the details of such a PES, the interaction should ideally be expanded over an infinite set of basis functions. However, only a limited number of expansion terms (~ 100) are needed for scattering calculations at low collision energy ($E_{coll} < 1000$ cm^{-1}). These terms correspond to projections of the PES on the basis functions of highest angular symmetry. If the potential energy could be easily estimated for any arbitrary orientation of the partners, it would be straightforward to compute the expansion terms by standard quadrature formulae or by a least squares fit of the PES on a truncated expansion of basis functions. However, the computation of a single potential energy point at high accuracy, i.e. with a precision ~ 1 cm^{-1}, is usually computer time consuming and the PES can be obtained only for a reduced set of orientations. For atom-molecule interactions which depend on only two angles (orientation of the intermolecular vector with respect to the rigid molecule frame), the main expansion terms can be obtained using a reasonable number of PES data points, of the order of 10–100 independent orientations for each intermolecular distance. For molecule-molecule interactions the number of PES data points can rise to thousands of orientations. Sheldon Green [4] has shown how slow the convergence of Gauss quadratures can be when there are as many as four angular degrees of freedom. From a general point of view, the difficulty is to solve an inverse problem numerically ill conditioned. We note that methods based on the singular value decomposition (SVD) have been developed in similar contexts, e.g. the correction of an existing PES using experimental spectroscopic data, see [5].

The present work focuses on optimizing the number and the distribution of PES data points needed for a given expansion basis set. In the case of a random distribution, we show that is is possible to estimate, using the Monte Carlo theorem, the accuracy on each expansion term. The general method and equations are derived in Sect. 2. An application to the $H_2O–H_2$ system is presented in Sect. 3. Conclusions are drawn in Sect. 4.

2 General method

2.1 Angular expansion of the interaction potential

The interaction potential energy between two rigid molecules, can be expanded over a given set of angular functions as follows:

$$V(R; \hat{r}) = \sum_i v_i(R) t_i(\hat{r}) \tag{1}$$

where R is the intermolecular distance and $\hat{r}$ designates any angular coordinates on the hypersurface. Practically for molecular interactions $\hat{r}$ defines the relative orientation

angles between the two molecules. In what follows the dependence on R will be dropped since we focus only on the angular dependence of the PES.

We consider $(t_i(\hat{r}))_{(i=1;\infty)}$ is a complete set of orthonormal basis functions. For a given collision energy only a restricted set of n expansion coefficients $v_{i=1,n}$ are needed.

The exact theoretical potential can be expanded on the reduced basis set of n functions $(t_i(\hat{r}))_{(i=1;n)}$ as follows:

$$V(\hat{r}) = \sum_{j=1}^{n} v_j t_j(\hat{r}) + R_n^{th}(\hat{r}) \tag{2}$$

$R_n^{th}(\hat{r})$ is the theoretical residual function which is orthogonal to all basis functions $(t_i(\hat{r}))_{(i=1;n)}$ selected in the expansion.

The exact expansion coefficients can be expressed as:

$$v_i = \int t_i(\hat{r}) V(\hat{r}) d\hat{r} = \sum_{j=1,n} \delta_{ij} \int t_j(\hat{r})(V(\hat{r}) - R^{th}(\hat{r})) d\hat{r} \tag{3}$$

In practice ab initio PES calculations can only be perfomed on a predefined set of N orientations of the molecules $\hat{r}_k$ ($k = 1, ..N$). The expansion coefficients v_i are derived from this potential surface sampling. From the potential expansion on each node $\hat{r}_{k=1,N}$ we derive a set of N equations with n unknowns v_j.

$$V(\hat{r}_k) = \sum_{j=1}^{n} v_j t_j(\hat{r}_k) + R_n^{th}(\hat{r}_k) \tag{4}$$

The v_i coefficients can be obtained by inversion of system (4), provided that the nodes grid size is large enough $N > n$. We can define a scalar product of two functions f and g known on the $N(\hat{r}_k)$ grid nodes as:

$$\langle f | g \rangle_N = \sum_{k=1,N} w_k f(\hat{r}_k) g(\hat{r}_k)$$

$\langle f | g \rangle_N$ is a quadrature approximation of the integral $\int f(\hat{r}) g(\hat{r}) d\hat{r}$ on the nodes $\hat{r}_k$ with weights w_k. From this scalar product one can obtain an overlap matrix S^N which gives the basis functions scalar products:

$$S_{ij}^N = \langle t_i | t_j \rangle_N$$

The exact potential coefficients v_i can be expressed as a function of the residual potential and the overlap S matrix.

$$v_i = S_{ij}^{-1} \left(\langle t_j | V \rangle_N - \langle t_j | R_n^{th} \rangle_N \right) \tag{5}$$

2.2 Least square fit approximation and optimisation of the angular sampling

Different fitting procedures are commonly used based on different approximations of the residual potential R^{th} and of the overlap matrix S_{ij}. In the least square fit approximation the potential terms v_i are approximated by α_i:

$$V(\hat{r}) = \sum_{i=1,n} \alpha_i t_i(\hat{r}) + R^N(\hat{r})$$

$R^N(\hat{r})$ is the potential residual which minimises the residual mean square (*rms*). Thus, $R^N(\hat{r})$ is orthogonal to all basis functions $\langle t_i | R^N \rangle_N = 0$ and α_i can be expressed as

$$\alpha_i = \sum_{j=1,n} S_{ij}^{-1} \langle t_j | V \rangle_N$$

The error on these coefficients is then

$$e_i = \alpha_i - v_i = \sum_{j=1,n} S_{ij}^{-1} \langle t_j | R_n^{th} \rangle_N \tag{6}$$

These coefficients are a good approximation for the v_i provided the theoretical residual function is orthogonal to all t_j basis functions. If the residual function can not be neglected, the potential fit is usually validated by *rms* deviation between the exact interaction energy and the approximated potential expansion at the nodes. In the case of least square fit, a too small sampling of the PES often reduces the *rms* at the cost of spurious non physical oscillations of the fit. In particular these oscillations occur for an ill-behaved sampling of the expansion basis functions on the PES. The overlap matrix S_{ij} (ideally close to δ_{ij}) is then ill conditionned. Therefore we suggest to test before hand (prior to any PES computation) the quality of the PES sampling for a chosen set of expansion basis functions. A simple minimal criterium is to ensure that $||S^{-1}|| \sim 1$.

However, even for an good sampling of all basis functions of a chosen expansion basis $t_{i=1,n}$, such as $S_{ij} = \delta_{ij}$, the potential residual can affect the expansion coefficients. Equation (6) shows how the residual R_n^{th} contributes to the coefficient error e_i. Unfortunately the potential residual can only be estimated after hand from the fit *rms*. The other alternative to reduce the error e_i is to extend the expansion basis size $t_{i=1,n'}$ so that the residual function should be a small correction to the analytical expansion. A new PES node grid ($\hat{r}_{k=1,N'}$) might be necessary to ensure a good sampling of all added basis functions. This procedure may be iterated.

Nevertheless, the number of data grows rapidly as the surface sampling is extended. Fitting procedure should be extensive on node distribution. Quadratures approximations based on orthogonal polynomials (such as Gauss quadratures) require one specific node distribution for each chosen potential fit basis size n. On the contrary quadratures based on irregular node distribution (as random quadratures) are extensive. Overall for many dimension problems irregular node distributions are necessary as any tensorial product of one dimension grid grows exponentially with the number of degrees of

freedom. Among all irregular distributions we consider in the present work random node grids. Random grids allow to estimate the fit error from the $1/\sqrt{N}$ convergence of random quadratures. We discuss in the following section the error estimation of a least square fit based on a random surface sampling.

2.3 Expansion error estimator for random samplings

For any random sampling all nodes are equivalent and the weight (w_k) is uniform over the whole surface $w_k = \tau/N = \int d\hat{r}/N$. The Monte Carlo theorem gives a statistical estimate of the random quadratures $\langle t_i | R_n^{th} \rangle_N$ [6]. In this section we show how to estimate the expansion term error as a function of the potential residual mean square and the overlap matrix S of the basis functions on the PES grid.

The least square fit error can be written as:

$$e_{\alpha i} = S_{ij}^{-1} \langle t_j | R_n^{th} \rangle_N = S_{ij}^{-1} e_i'$$

Each coefficient e_i' is the random quadrature of the $(t_i R_i^{th})$ function. Therefore the Monte Carlo theorem gives a statistical estimate of e_i' proportional to $1/N^\alpha$ with $\alpha = 0.5$ for a purely random set of points.

$$e_i' = \sum_k \frac{\tau}{N} t_i(\hat{r}_k) R_n^{th}(\hat{r}_k)$$

$$\sim \int t_i(\hat{r}) R_n^{th}(\hat{r}) d\tau \pm \frac{1}{N^\alpha} \sqrt{\frac{\tau^2}{N} \sum_k t_i^2(\hat{r}_k) R_n^{th^2}(\hat{r}_k) - \left(\sum_k \frac{\tau}{N} t_i(\hat{r}_k) R_n^{th}(\hat{r}_k)\right)^2}$$

$$\sim \int t_i(\hat{r}) R_n^{th}(\hat{r}) d\tau \pm \frac{1}{N^\alpha} \sqrt{\frac{\tau^2}{N} \sum_k t_i^2(\hat{r}_k) R_n^{th^2}(\hat{r}_k) - e_i'^2}$$

An overestimation of $t_i^2(\hat{r}_k)$ is given by $T_i^2 \sim Max_k |t_i^2(\hat{r}_k)|$.

$$e_i' \sim \frac{1}{N^\alpha} \sqrt{T_i^2 \frac{\tau^2}{N} \sum_k R_n^{th^2}(\hat{r}_k) - e_i'^2}$$

$$e_i' \sim \frac{1}{N^\alpha} \sqrt{T_i^2 \ \tau^2 rms^{th^2} - e_i'^2}$$

rms^{th} is the theoretical residual mean square of the exact potential fit. We show below how it depends on the approximate least square fit rms, on the overlap matrix S and on the error coefficients e_i.

$$rms^{th^2} = \frac{1}{N} \sum_{k=1}^{N} R_n^{th^2}(\hat{r}_k)$$

$$= \frac{1}{N} \sum_{k=1}^{N} \left(R^N(\hat{r}_k) - \sum_{j=1}^{n} e_j t_j(\hat{r}_k) \right)^2$$

$$= rms^2 + \frac{1}{\tau} \sum_{j,j'=1}^{n} e_j S_{jj'} e_{j'}$$

$$= rms^2 + \frac{1}{\tau} \sum_{j,j'=1}^{n} e_j' S_{jj'}^{-1} e_{j'}'$$

We obtain an estimate of each e_i' coefficient:

$$e_i'^2 \sim \frac{1}{N^{2\alpha}} \left(T_i^2 \left(\tau^2 \, rms^2 + \tau \sum_{jj'} e_j' S_{jj'}^{-1} e_{j'}' \right) - e_i'^2 \right)$$

This expression couples the error e_i' to all other coefficients e_j'. Therefore we choose to estimate the error coefficients root mean square or the error norm $||e'||_\infty^2 = \sum_i e_i'^2$:

$$||e'||_\infty^2 \leq \frac{1}{N^{2\alpha}} \left(\sum_{i=1}^{n} T_i^2 \left(\tau^2 \, rms^2 + \tau \sum_{jj'} e_j' S_{jj'}^{-1} e_{j'}' \right) - \sum_{i=1}^{n} e_i'^2 \right)$$

$$\leq \frac{1}{N^{2\alpha}} \left(\sum_{i=1}^{n} T_i^2 \left(\tau^2 \, rms^2 + \tau < e'|S^{-1}e' >_\infty \right) - \sum_{i=1}^{n} e_i'^2 \right)$$

To simplify the expression we define T as

$$T^2 = \frac{1}{n} \sum_{i=1}^{n} \text{Max}_k \left(t_i^2(\hat{r}_k) \right) = \frac{1}{n} \sum_{i=1}^{n} T_i^2$$

We obtain the algebric expression for $||e'||_\infty^2$:

$$||e'||_\infty^2 \leq \frac{1}{N^{2\alpha}} \left(nT^2 \left(\tau^2 \, rms^2 + \tau < e'|S^{-1}e' >_\infty \right) - ||e'||_\infty^2 \right) \tag{7}$$

The Schwarz theorem gives an overestimate of $< e'|S^{-1}e' >_\infty$

$$< e'|S^{-1}e' >_\infty \leq ||S^{-1}|| \, ||e'||_\infty^2$$

where the S^{-1} matrix norm is chosen as it's maximum eigenvalue:

$$||S^{-1}|| = \text{Max}_i |\lambda_i| \tag{8}$$

Thus

$$||e'||_\infty^2 \leq \frac{1}{N^{2\alpha}} \left(n \, T^2 \, \tau^2 \, rms^2 + ||e'||_\infty^2 (||S^{-1}|| \, n \, T^2 \, \tau - 1) \right)$$

and provided that

$$\rho(n, N) = \frac{|(||S^{-1}|| \, n \, T^2 \, \tau - 1)|}{N^{2\alpha}} < 1$$

we obtain

$$||e'||_\infty^2 \leq \frac{1}{N^{2\alpha}} n \, T^2 \, \tau^2 \, rms^2 \left(1 - \frac{1}{N^{2\alpha}} \left(||S^{-1}|| \, n \, T^2 \, \tau - 1 \right) \right)^{-1}$$

The two error functions $e(\hat{r})$ and $e'(\hat{r})$ are related by

$$e = S^{-1} e'$$
$$||e||_\infty^2 = ||S^{-1} e'||_\infty^2 \leq ||S^{-1}||^2 \, ||e'||_\infty^2$$

Finally:

$$||e||_\infty^2 \leq ||S^{-1}||^2 \frac{1}{N^{2\alpha}} n \, T^2 \, \tau^2 \, rms^2 \, (1 - \rho(n, N))^{-1}$$

We obtain an explicit upper limit of $||e||_\infty$

$$||e||_\infty \leq \frac{\sqrt{n}}{N^\alpha} \tau \, T \, rms \, ||S^{-1}|| \, (1 - \rho(n, N))^{-1/2}$$

If we consider that all basis function are normalized to unity we get an approximation of $T \sim 1/\sqrt{\tau}$ and a more practical error estimator:

$$||e||_\infty \sim \frac{\sqrt{n\tau}}{N^\alpha} rms \, ||S^{-1}|| \, (1 - \rho(n, N))^{-1/2}$$

The mean error on each expansion coefficient can be estimated as:

$$e_i \sim ||e||_\infty / \sqrt{n} \tag{9}$$
$$\sim \sqrt{\frac{\tau}{N}} rms \, ||S^{-1}|| \, (1 - \rho(n, N))^{-1/2} \tag{10}$$

with

$$\rho(n, N) = \frac{|(||S^{-1}|| \, n - 1)|}{N^{2\alpha}} < 1$$

and $||S^{-1}||$ is chosen as the maximum S^{-1} eigenvalue. We obtain a convergence criterium $\rho(n, N) < 1$ which depends only on the given set of expansion basis functions $(t_i)_n$ and on the chosen random grid nodes $(\hat{r}_k)_N$.

For an ideal set of angular geometries the overlap S matrix and its inverse S^{-1} are close to the unit matrix. An ill conditionned sampling of the basis functions is characterized by very low eigenvalues of the S matrix, very large eigenvalues and therefore a very large norm of the S^{-1} matrix. As a result, given an expansion set of n basis function, we can optimize the distribution of N angular geometries by monitoring the value of S^{-1} and $\rho(n, N)$ for any set of basis functions. The final error estimate, Eq. 9, is proportional to the residual mean square. Therefore for each set of N potential data there is an optimal expansion basis for which the *rms* is minimum while the S matrix remains well conditioned. For such expansion basis set $(t_i)_n$, the mean error on each expansion term is optimized.

3 Application to the H_2O–H_2 system

The H_2O–H_2 interaction provides a very interesting system to assess the interest of the diagnostic tools presented in the previous section. The H_2O–H_2 system is of particular interest in astrophysics as the observed intensities of the water lines in the dense interstellar medium are very strongly linked to the excitation of H_2O by H_2 [7]. As the computation of collisional excitation rate coefficients requires the determination of an intermolecular potential energy surface, the accuracy of the fitting procedure is a key ingredient of the theoretical evaluation.

3.1 Ab initio quantum chemical calculations

There have been several ab initio PES developed for H_2O–H_2 [8,9,3] and there also exists a number of experimental results which can be used to probe the accuracy of the available potential surfaces, see e.g. the calculations of the rovibrational states of the complex [10,11] or the molecular beam experiments of Belpassi et al. [12]. Dubernet et al. [13] also discussed the influence of the H_2O–H_2 PES on the rotational excitation of H_2O and have shown that PES effects are particularly important at low temperatures ($T < 20$ K).

The most accurate H_2O–H_2 intermolecular potential has been calculated by Valiron et al. [3] and has been used by several authors to compute rotational excitation rate coefficients [7,14–17]. The PES of Valiron et al. include nine degrees of freedom. In the present work, only the rigid-rotor, five dimensional, reference PES at the CCSD(T) level of theory will be discussed. This PES was computed for 25 intermolecular distances between 3 and 15 au and 3,000 random angular geometries for each distance. In the following, we will also discuss a set of CCSD(T) data points obtained at regular angular geometries. These points were not employed by Valiron et al. Full details on the monomer geometries and basis sets can be found in Valiron et al. [3].

A previous H_2O–H_2 PES was computed by Phillips et al. [8] and was also employed for rotational excitation studies [18,19]. Phillips et al. computed the intermolecular potential between H_2O and H_2 for a total of 722 geometries using MP4(SDTQ)

perturbation theory. The resulting points were fitted to a 48 term angular expansion. We note that when good convergence is obtained, it is expected that MP4(SDTQ) and CCSD(T) calculations are similar [20]. The accuracy of the reference PES of Valiron et al. [3] should be therefore comparable to the results obtained by Phillips et al. [8]. However, it is important to note that Phillips et al. [8] performed their global fit from *ab initio* calculations on a restricted regular angular grid (see below). They used a three step process based on interpolation to generate the potential at additional orientations, then interpolation at each orientation to get a finer radial grid, and finally a linear least squares fit to the expanded data set. It is therefore not possible to directly compare our procedure with their fit as their final grid is not known.

3.2 Expansion of the H_2O–H_2 interaction

The discussion below focuses on 2 intermolecular distances, namely $R = 6$ and 5 au. These 2 distances were chosen in order to investigate the short-range part of the H_2O–H_2 interaction which is highly anisotropic. We note that the global minimum of the reference (uncorrected) PES of Valiron et al. is at $R = 5.9$ au with a well depth of -209 cm^{-1}. We used the coordinate conventions of Phillips et al. [8] which define θ, ϕ and θ', ϕ' as the collision direction and the H_2 orientation relative to the water molecule body-fixed axis system. In terms of these relative coordinates, the intermolecular potential can be written as [8]

$$V(R, \theta, \phi, \theta', \phi') = \sum v_{l_1 m_1 l_2 l}(R) t_{l_1 m_1 l_2 l}(\theta, \phi, \theta', \phi') \tag{11}$$

with

$$t_{l_1 m_1 l_2 l}(\theta, \phi, \theta', \phi') = \alpha_{l_1 m_1 l_2 l}(1 + \delta_{m_1 0})^{-1} \sum \begin{pmatrix} l_1 & l_2 & l \\ r_1 & r_2 & r \end{pmatrix} Y_{l_2 r_2}(\theta', \phi') * Y_{lr}(\theta, \phi)$$
$$* \left[\delta_{m_1 r_1} + (-1)^{l_1 + m_1 + l_2 + l} \delta_{-m_1 r_1} \right]$$

where $\alpha_{l_1 m_1 l_2 l}$ is the normalization factor [3]:

$$\alpha_{l_1 m_1 l_2 l} = [2(1 + \delta_{m_1 0})^{-1}(2l_1 + 1)^{-1}]^{-1/2} \tag{12}$$

The integer indices l_1, m_1, l_2 and l refer to the tensor ranks on the angle dependence of the H_2O orientation, the H_2 orientation and the collision vector orientation, respectively. The sum is over r_1, r_2, r. As discussed by Phillips et al. [8], the symmetry with respect to the reflection of H_2 in the H_2O plane is explicitly taken into account by the phased sum over $\pm m_1$. In the rigid-rotor approximation, the C_{2v} symmetry of H_2O further requires that m_1 be even. The homonuclear symmetry of H_2 similarly constrains l_2 to be even. It should be noted that the correspondence between the expansion angular coefficients $v_{l_1 m_1 l_2 l}(R)$ and their counterpart in terms of the electrostatic, induction and dispersion contributions has been given by Stone et al. [21].

In order to investigate the influence of the angular grid, we used the same set of angular basis functions as Phillips et al [8]. These include the symmetry-allowed terms

Table 1 Angular coefficients $v_{l_1 m_1 l_2 l}$ (in cm^{-1}) at a H$_2$O–H$_2$ intermolecular separation $R = 6$ au

$l_1 m_1 l_2 l$	$v_{l_1 m_1 l_2 l}(R)$[a]	$v_{l_1 m_1 l_2 l}(R)$[b]	$v_{l_1 m_1 l_2 l}(R)$[c]
0000	−376.95	−345.34	−343.17
1001	114.30	131.59	116.31
1023	704.66	726.29	728.14
2202	−168.00	−127.09	−147.76
2224	542.25	533.23	591.77
3025	−102.40	−86.206	−121.60
3225	209.03	209.88	215.72
4226	64.579	64.691	59.794
A	2618.8	2773.2	2380.8
R	713.57	894.24	399.74

[a] Fit obtained by Phillips et al. [8], [b] fit obtained from our CCSD(T) calculations with a (small) angular grid of 75 angular geometries taken from Phillips et al. [8], [c] fit obtained from our CCSD(T) calculations with a (large) random grid of 3,000 geometries. Only the 8 largest coefficients are listed. The parameter A and R gives the sum of the absolute values of, respectively, the anisotropic coefficients (all terms except $v_{0000}(R)$) and the 40 remaining (not listed) coefficients. See text

for $l_1 \leq 6$, $m_1 \leq 4$, $l_2 \leq 2$, and $l \leq 8$ except for terms with $l = l_1$, $l_2 = 2$, which were included only for $l_1 \leq 3$, $m_1 \leq 2$. This selection provide 48 angular coefficients at each distance. The 8 largest angular coefficients obtained by Phillips et al. for $R = 6$ and 5 au are presented in Tables 1 and 2. The regular grid used by Phillips et al. consists of 75 and 83 angular geometries for $R = 6$ and 5 au, respectively.

As a consequence, we carried out CCSD(T) calculations on two different angular grids: a small one which consists of the 75 and 83 angular geometries published by Phillips et al. [8] and a large one which consist of 3,000 random geometries for both $R = 6$ and 5 au. **The employed random generator provides a uniform distribution of $cos\theta (cos\theta')$ and $\phi(\phi')$, respectively on $[-1, 1]$ and $[0, 2\pi]$, consistent with the requirement of a uniform sampling of the differential solid angle $sin\theta d\theta d\phi (sin\theta' d\theta' d\phi')$. We note that the symmetries of the system are included in the chosen potential basis functions. Therefore, since the sampling is random, there is no need to restrict the *ab initio* geometries according to the molecular symmetries.** The same 3,000 random geometries were used in Valiron et al. [3].

Results are presented in Tables 1 and 2 along with the results of Phillips et al. [8]. It can be observed that the choice of the grid does significantly affect some of the listed coefficients of the fit. In particular, at $R = 6$ au, $|v_{3025}(R)|$ shows an increase of 41 percent between the small and the large grid. We can also notice that the global anisotropy, as given by the parameter A, is largely reduced when using the large grid. Indeed this can be explained by summing the absolute values of the 40 smallest (not listed) coefficients, as given by the parameter R: we observe that using the small grid gives more weight to highly anisotropic terms, in particular at $R = 6$ au where the potential is intrinsically less anisotropic than at 5 au. This effect is also observed in the fit obtained by Phillips et al. [8] but to a lesser extent as a result of the three step procedure used by these authors.

Table 2 Angular coefficients $v_{l_1m_1l_2l}$ (in cm^{-1}) at a H$_2$O–H$_2$ intermolecular separation $R = 5$ au	$l_1m_1l_2l$	$v_{l_1m_1l_2l}^{\text{a}}$	$v_{l_1m_1l_2l}^{\text{b}}$	$v_{l_1m_1l_2l}^{\text{c}}$
	0000	5449.8	5485.8	5454.7
	0022	566.09	588.06	571.03
	1001	1540.3	1618.1	1563.9
	1023	1632.5	1658.6	1658.6
	2224	1587.0	1611.1	1595.2
	3003	−571.38	−587.50	−545.78
Same notations as in Table 1 except $^{\text{b}}$fit obtained from our CCSD(T) calculations with an angular grid of 83 angular geometries taken from Phillips et al. [8]	3203	1026.8	1057.2	1019.2
	3225	771.03	766.72	743.84
	A	10851	11645	10579
	R	3155.6	3758.1	2881.5

The spurious increase of anisotropy on the small angular grids was in fact expected since the least squares fit procedure must compensate a poor sampling of the angular basis functions by introducing numerical artefacts. As discussed in Sect. 2.2, a well suited angular sampling of orientations should give a good estimation of the scalar products of the basis functions. A poor sampling of the basis functions is then characterized by very large eigenvalues of the inverse S^{-1} matrix. The norm $||S^{-1}||$ therefore provides an interesting probe of the insufficiencies of the grid and of the numerical instabilities induced in the least squares fit procedure.

This point is illustrated in Fig. 1 which plots the convergence of the norm of the inverse S^{-1} matrix with the number of geometries. It is shown that a random angular grid with less than 100 geometries gives a $||S^{-1}||$ larger than 10. About 1,000 geometries are needed to obtain a $||S^{-1}||$ lower than 2. It should be noted that the 75 regular geometries taken from Phillips et al. [8] gives a $||S^{-1}||$ of 358! By contrast, 3,000 random geometries provide a $||S^{-1}||$ of 1.40, indicating the reliability of the fit employed by Valiron et al. [3]. We note also that beyond 3,000 geometries, the convergence is slow and a modest lowering of $||S^{-1}||$ would require a significant additional number of geometries.

The efficiency of a random sampling is also illustrated in Fig. 2. We investigated the convergence of $||S^{-1}||$ by adding random orientations to the 75 regular geometries of Phillips et al. [8]. As can be observed, $||S^{-1}||$ is decreased from 358 down to 13.7 by using only 20 additional random geometries. This shows the fundamental advantage of using extensive distribution of geometries, as discussed in Sect. 2.2.

Finally, the relevance of the error estimator introduced in Sect. 2.3 is illustrated in Fig. 3 for the two leading angular coefficients v_{0000} and v_{1023} at $R = 5$ and 6 au. It is observed that the error on these expansion coefficients is well predicted by our estimator (Eq. 9), and that a convergence better than ∼1% is obtained as soon as the number of random geometries exceeds ∼1,000.

It should be noted that, in practice, the selection of the final set of basis functions can be obtained iteratively, as done by Valiron et al. [3] for H$_2$O–H$_2$. These authors first selected a maximal expansion that included all anisotropies up to $l_1 = 12$ for H$_2$O
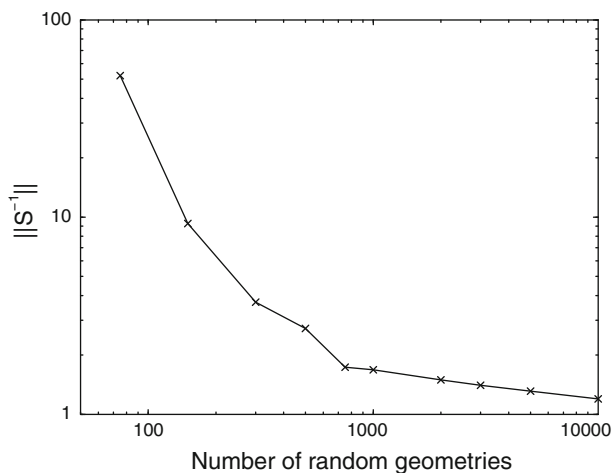
Fig. 1 Convergence of the norm of the inverse S^{-1} matrix as a function of the number of random geometries
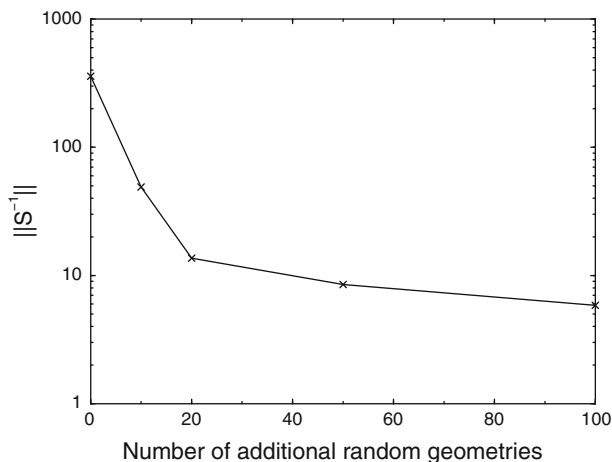

Fig. 2 Convergence of the norm of the inverse S^{-1} matrix as a function of the number of random geometries added to the 75 regular geometries taken from Phillips et al. [8]

and $l_2 = 6$ for H_2, resulting in 1150 $t_{l_1 m_1 l_2 l}$ functions. They then selected iteratively all statistically significant terms using the following procedure at an intermolecular distance of $R = 5$ au: they started the fit with a minimal expansion limited to the single t_{0000} function and evaluated the 1,149 remaining terms by random quadratures of $< t_j | V >_N$. All terms above $4 \times e_i$, where e_i is the mean error as defined in Eq. 9, were then selected and added to t_{0000} providing a new starting expansion. This process was iterated until convergence. The final set was composed of 149 angular functions including anisotropies up to $l_1 = 11$ and $l_2 = 6$. We note that a cubic spline (radial) interpolation of each expansion term was finally employed over the whole intermolecular distance range and was smoothly connected with standard extrapolations to
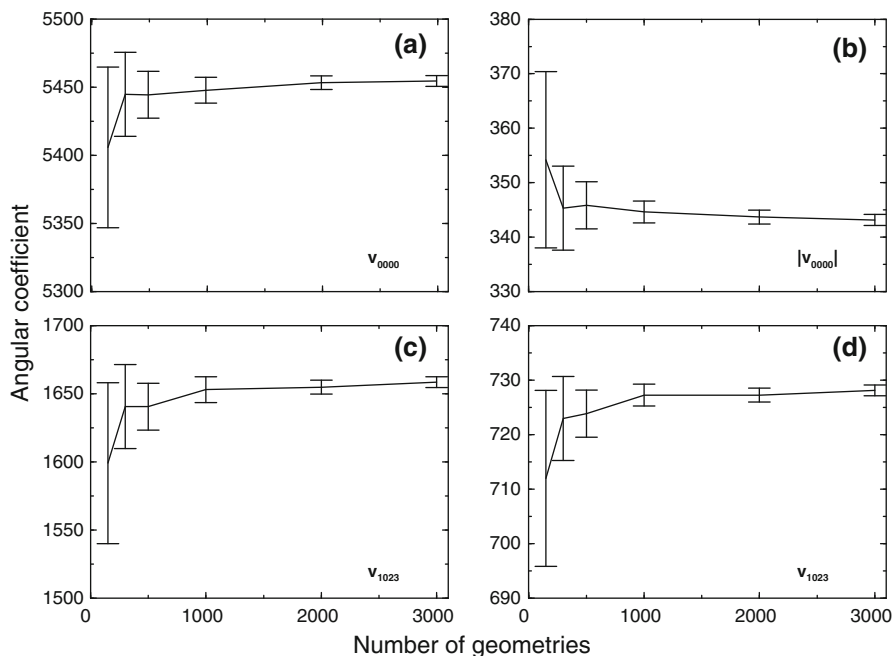
Fig. 3 Convergence of the error estimator as a function of the number of (random) geometries for the two leading angular coefficients v_{0000} and v_{1023}: **a** v_{0000} at $R = 5$ au, **b** $|v_{0000}|$ at $R = 6$ au, **c** v_{1023} at $R = 5$ au, **d** v_{1023} at $R = 6$ au

provide continuous radial expansion coefficients suitable for scattering calculations, as explained in Valiron et al. [3].

4 Conclusion

We have investigated in this paper how to monitor the accuracy of the angular expansion of rigid-rotor PES for molecule-molecule systems. We have shown that the use of random grids for the sampling of the angular geometries combined with the Monte Carlo theorem allows to estimate the mean accuracy on each expansion term determined by a least squares fit. The interest of our approach is illustrated by an application to the H_2O–H_2 system, of great astrophysical interest. We have found that the inverse of the S matrix, which gives the scalar products of the basis functions, provides a useful probe of the insufficiencies of the grid and of the related numerical instabilities induced by the least squares fit. We have also shown the advantage of using extensive random grid which should contain more than ∼1,000 data points to ensure a convergence better than ∼1% for the leading terms of the angular expansion. To conclude, we note that our procedure was employed recently for a number of molecule-H_2 systems of astrophysical interest: CO–H_2 [22], NH_3–H_2 [23], H_2CO–H_2 [24], SO_2–H_2 [25], HDO–H_2 [17] and ND_2H–H_2 [26].

References

1. T.R. Phillips, S. Maluendes, S. Green, J. Chem. Phys. **102**, 6024 (1995)
2. A. Faure, P. Valiron, M. Wernli, L. Wiesenfeld, C. Rist, J. Noga, J. Tennyson, J. Chem. Phys. **122**, 221102 (2005)
3. P. Valiron, M. Wernli, A. Faure, L. Wiesenfeld, C. Rist, S. Kedzuch, J. Noga, J. Chem. Phys. **129**, 134306 (2008)
4. S. Green, J. Chem. Phys. **103**, 1035 (1995)
5. Q. Wu, J.Z.H. Zhang, J.M. Bowman, J. Chem. Phys. **107**, 3602 (1997)
6. A.H. Stroud, Approximate Calculation of Multiple Integrals, Prentice-Hall, Inc (1971)
7. A. Faure, N. Crimier, C. Ceccarelli, P. Valiron, L. Wiesenfeld, M.L. Dubernet, Astron. Astrophys. **472**, 1029 (2007)
8. T.R. Phillips, S. Maluendes, A.D. McLean, S. Green, J. Chem. Phys. **101**, 5824 (1994)
9. M.P. Hodges, R.J. Wheatley, G.K. Schenter, A.H. Harvey, J. Chem. Phys. **120**, 710 (2004)
10. A. van der Avoird, D.J. Nesbitt, J. Chem. Phys. **134**, 044314 (2011)
11. X.-G. Wang, T. Carrington, J. Chem. Phys. **134**, 044313 (2011)
12. L. Belpassi, M.-L. Reca, F. Tarantelli, L.-F. Roncaratti, F. Pirani, D. Cappelletti, A. Faure, Y. Scribano, J. Am. Chem. Soc. **132**, 13046 (2010)
13. M.L. Dubernet, F. Daniel, A. Grosjean, A. Faure, P. Valiron, M. Wernli, L. Wiesenfeld, C. Rist, J. Noga, J. Tennyson, Astron. Astrophys. **460**, 323 (2006)
14. M.L. Dubernet, F. Daniel, A. Grosjean, C.Y. Lin, Astron. Astrophys. **497**, 911 (2009)
15. F. Daniel, M.L. Dubernet, F. Pacaud, A. Grosjean, Astron. Astrophys. **517**, A13 (2010)
16. Y. Scribano, A. Faure, L. Wiesenfeld, J. Chem. Phys. **133**, 231105 (2010)
17. L. Wiesenfeld, Y. Scribano, A. Faure, Phys. Chem. Chem. Phys. **13**, 8230 (2011)
18. T.R. Phillips, S. Maluendes, S. Green, Astrophys. J. **107**, 467 (1996)
19. M.L. Dubernet, A. Grosjean, Astron. Astrophys. **390**, 793 (2002)
20. R.J. Bartlett, Ann. Rev. Phys. Chem. **32**, 359 (1981)
21. A.J. Stone, R.J.A. Tough, Chem. Phys. Lett. **110**(2), 123 (1984)
22. M. Wernli, P. Valiron, A. Faure, L. Wiesenfeld, P. Jankowski, K. Szalewicz, Astron. Astrophys. **446**, 367 (2006)
23. S. Maret, A. Faure, E. Scifoni, L. Wiesenfeld, Mon. Not. R. Astron. Soc. **399**, 425 (2009)
24. N. Troscompt, A. Faure, L. Wiesenfeld, C. Ceccarelli, P. Valiron, Astron. Astrophys. **493**, 687 (2009)
25. A. Spielfiedel, M.-L. Senent, F. Dayou, C. Balança, L. Cressiot-Vincent, A. Faure, L. Wiesenfeld, N. Feautrier, J. Chem. Phys. **131**, 014305 (2009)
26. L. Wiesenfeld, E. Scifoni, A. Faure, E. Roueff, Mon. Not. R. Astron. Soc. **413**, 509 (2011)