The Journal of Chemical Physics

Neural network based quasi-diabatic Hamiltonians with symmetry adaptation and a correct description of conical intersections

Cite as: J. Chem. Phys. **150**, 214101 (2019); https://doi.org/10.1063/1.5099106
Submitted: 06 April 2019 . Accepted: 10 May 2019 . Published Online: 03 June 2019

Yafu Guan, Hua Guo, and David R. Yarkony

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Neural network based quasi-diabatic Hamiltonians with symmetry adaptation and a correct description of conical intersections

Cite as: J. Chem. Phys. 150, 214101 (2019); doi: 10.1063/1.5099106
Submitted: 6 April 2019 • Accepted: 10 May 2019 •
Published Online: 3 June 2019

Yafu Guan,[1,a)] Hua Guo,[2,b)] and David R. Yarkony[1,c)]

AFFILIATIONS

[1] Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, USA
[2] Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

[a)] Electronic mail: yguan15@jhu.edu
[b)] Electronic mail: hguo@unm.edu
[c)] Electronic mail: yarkony@jhu.edu

ABSTRACT

In a previous paper, we have demonstrated that artificial neural networks (NNs) can be used to generate quasidiabatic Hamiltonians ($\mathbf{H}^{d}$) that are capable of representing adiabatic energies, energy gradients, and derivative couplings. In this work, two additional issues are addressed. First, symmetry-adapted functions such as permutation invariant polynomials are introduced to account for complete nuclear permutation inversion symmetry. Second, a partially diagonalized representation is introduced to facilitate a better description of near degeneracy points. The diabatization of 1, $2^{1}A$ states of NH_3 is used as an example. The NN fitting results are compared to that of a previous fitting with symmetry adapted polynomials.

I. INTRODUCTION

Accurate quantum dynamics simulations of electronically nonadiabatic processes involving conical intersections require accurate analytic representations of the requisite electronic structure data, energies, energy gradients, and derivative couplings in the adiabatic representation.[1,2] However, as the nuclear coordinates approach a conical intersection where a degeneracy occurs, the derivative coupling becomes singular, which causes the nuclear wave functions for different electronic states to be strongly coupled. In addition, the corresponding adiabatic potential energy surfaces (PESs) have discontinuous derivatives at the degeneracy. Nuclear dynamics in the adiabatic basis is difficult to implement due to this singularity in the derivative coupling and the nonsmooth potentials. The difficulties of adiabatic representation motivate the use of quasidiabatic representations. In a diabatic representation, the singular derivative coupling vanishes globally, which simplifies the form of the nuclear Schrödinger equation. For polyatomic molecules, rigorous diabatic bases do not exist;[3–5] hence, the attribute *quasi* is used above, and will be omitted below except as needed for emphasis.

The construction of a quasidiabatic representation for polyatomic chemical system has long been a challenging task in computational nonadiabatic chemistry. Currently available methods fall into four categories: property-based methods,[6,7] methods based on configuration uniformity,[8,9] diabatization by ansatz,[10,11] and coupling-based methods. The coupling-based methods are in principle the most accurate as they use the derivative coupling in diabatizing the electronic states. With the information of derivative coupling, the quality of diabatization can be strictly under control; otherwise, it would depend to a large extent on the physical intuition for a specific system.[12] Some coupling-based methods reported in the literature include (i) solving the Poisson equation,[13–15] (ii) the Shepard interpolation,[16–18] and (iii) fitting using symmetry adapted polynomials, as demonstrated by Zhu and Yarkony (ZY).[19–22] In ZY fitting, the diabatic Hamiltonian is expressed in symmetry adapted polynomials and the procedure simultaneously fits and diabatizes the *ab initio* electronic structure

data obtained in the adiabatic representation to produce a coupled quasidiabatic representation. As this approach uses derivative coupling data to determine diabatic states, the residual derivative coupling can be determined and used to assess the quality of the diabatization. This approach ultimately provides an accurate, quantifiably quasidiabatic representation in a least squares sense. Its accuracy has been demonstrated by excellent agreement with experimentally measured dynamical attributes in photodissociation of ammonia.[23–26]

For years, Artificial Neural Networks (NNs) have been used as a robust and powerful tool to fit accurate adiabatic PESs for polyatomic molecules in the gas phase and for the interaction of small molecules with metal surfaces.[27–41] Moreover, an NN also provides a closed analytic form for gradients, and it has been used for simultaneous fitting of energies and energy gradients.[42,43] Therefore, an NN can be used as a coupling-based method to construct a diabatic representation. The flexible functional form of an NN can in principle represent any real and smooth function to any accuracy.[44] The removal of discontinuities in the derivative coupling near a conical intersection region in a diabatic representation also facilitates the fitting of NN-based quasidiabatic Hamiltonians. Early attempts have been made to construct diabatic representations with NNs.[10,45,46] We have also proposed a general fitting procedure with NNs based on the ZY diabatization method, which simultaneously fits adiabatic energies, energy gradients, and derivative couplings in a least squares sense.[47] Using a simple test example LiFH, the NN fitting has been demonstrated to accurately reproduce energies, energy gradients, and derivative couplings, and most importantly, it also allows the reproduction of quantum dynamic results generated from the original quasidiabatic representation, which indicates its global accuracy. However, two important issues remain to be addressed: One is the description of the vicinity of a conical intersection seam, and the other is the symmetry adaptation to account for the complete nuclear permutation inversion (CNPI) symmetry.

In this work, we will address these two issues in constructing a NN-based diabatic Hamiltonian for 1, 2^1A states of NH_3, which has been used in previous work on quasidiabatic representations and provides an archetypical test example. The 1, 2^1A states are coupled by a conical intersection seam, which plays a central role in the nonadiabatic photodissociation of ammonia,

$$NH_3(\tilde{X}, v) \rightarrow NH_3(\tilde{A}, v') \rightarrow NH_2(\tilde{X}, \tilde{A}) + H. \tag{1}$$

Three hydrogen atoms of ammonia are permutable, thus giving rise to the CNPI group $\mathbf{G}_{12}$ (isomorphic to $S_3 \otimes I$, direct product of the symmetric group of order 3 and the inversion group).[48] First, symmetry-adapted functions such as permutation invariant polynomials (PIPs)[49,50] are introduced to account for the CNPI symmetry. Second, a partially diagonalized representation[22] is used to facilitate a better description of near degeneracy points, which allows a correct description of the vicinity of a conical intersection seam.

The paper is organized as follows: Sec. II reports the construction of an NN-based quasidiabatic Hamiltonian; in Sec. III, the fitting results for NH_3 are presented. Section IV summarizes and suggests future directions to be pursued.

II. THE DIABATIC HAMILTONIAN AND ITS DETERMINATION WITH NEURAL NETWORKS

A. The quasidiabatic Hamiltonian

The diabatic Hamiltonian $\mathbf{H}^d$ is the matrix representation of the electronic Hamiltonian $\mathbf{H}^e(\mathbf{q}; \mathbf{R})$ in diabatic electronic basis $(\mathbf{\Psi}^{\mathrm{d}}_\alpha, \alpha = 1 - N^{\mathrm{state}})$

$$H^d_{\alpha,\beta}(\mathbf{R}) \equiv \langle \mathbf{\Psi}^d_\alpha(\mathbf{q};\mathbf{R})|\mathbf{H}^e(\mathbf{q};\mathbf{R})|\mathbf{\Psi}^d_\beta(\mathbf{q};\mathbf{R})\rangle_{\mathbf{q}}, \tag{2}$$

where $\mathbf{q}$ are the electronic coordinates and $\mathbf{R}$ are the internal coordinates that describe molecular geometries. The eigenvectors of $\mathbf{H}^d$ satisfy the following electronic Schrödinger equation:

$$[\mathbf{H}^d(\mathbf{R}) - \mathbf{I}E^{a,J,(m)}(\mathbf{R})]\mathbf{d}^J(\mathbf{R}) = \mathbf{0}, \tag{3}$$

where $\mathbf{I}$ is the identity matrix and $E^{a,J,(m)}$ is the corresponding eigenenergy. The superscript (m) indicates that the results come from the model Hamiltonian $\mathbf{H}^d$, rather than *ab initio* (*ab*) calculations, and the superscript (a) indicates the adiabatic representation. $\mathbf{H}^d$ is also called a potential energy matrix (PEM).

B. Feed-forward neural networks

The matrix elements of a PEM are continuous functions of internal coordinates. In this work, they are represented by feed-forward NN functions.

A multilayer feed-forward NN consists of an input layer, one or more hidden layers, and an output layer, which can be denoted as R–S^1–S^2–...–S^M. It means that the network has R elements in the input vector and S^m neurons in layer m. In a feed-forward NN, the output of one layer becomes the input of the following layer. The information flows through each layer in only one direction, and there are no feedback connections in which outputs are fed back into itself,[51] which makes the feed-forward NN the simplest type of artificial neural networks. The input layer is denoted as the zeroth layer ($S^0 = R$). The equations that describe the connection between layers are

$$n^m_i = \sum_{j=1}^{S^{m-1}} (w^m_{i,j} a^{m-1}_j) + b^m_i, \tag{4}$$

$$a^m_i = f^m(n^m_i), \tag{5}$$

where a^m_i denote the output of neuron i in layer m, n^m_i is the net input for layer m, f^m is the transfer function, $w^m_{i,j}$ represents the element at row i and column j of the weight matrix coupling layer $m-1$ to layer m, and b^m_i denotes the bias of neuron i in layer m. Figure 1 is an example of a feed-forward NN with two hidden layers. The number of adjustable parameters $\boldsymbol{\theta}$ (weights and biases) in such an NN is given by

$$N_\theta = \sum_{m=1}^{M} (S^{m-1} + 1) S^m. \tag{6}$$

Feed-forward NNs also provide a closed analytical form for the gradient, thus facilitating a coupling-based diabatization. The expressions for gradients and derivatives with respect to NN parameters have been reported previously.[47]

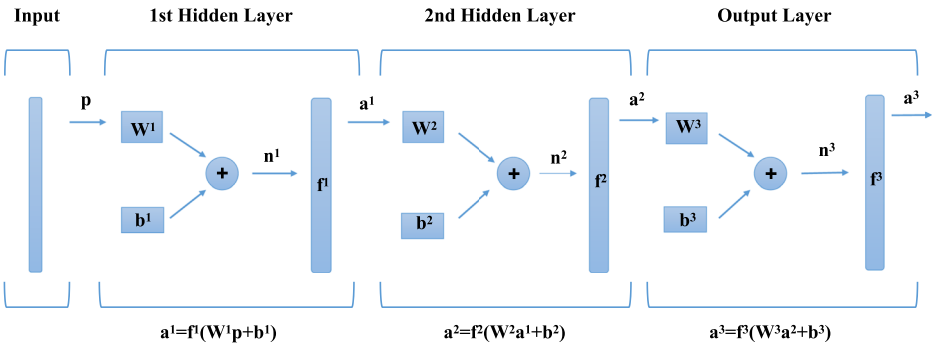

FIG. 1. A feed-forward neural network with two hidden layers.

C. Determination of $\mathbf{H}^d$ from *ab initio* data

The nonlinear parameters $\boldsymbol{\theta}$ (weights and biases) in NN are optimized by minimizing the differences between $\mathbf{H}^d$ determined and *ab initio* determined energies, energy gradients, and interstate couplings. For $x = ab$, define

$$L_0^{I,I,(x)}(\mathbf{R}) \equiv E^{a,I,(x)}(\mathbf{R}), \tag{7}$$

$$L_j^{I,I,(x)}(\mathbf{R}) \equiv \nabla_j E^{a,I,(x)}(\mathbf{R}), \tag{8}$$

$$L_j^{I,J,(x)}(\mathbf{R}) \equiv h_j^{a,I,J,(x)}(\mathbf{R}), \tag{9}$$

where $h_j^{a,I,J,(x)}$ is the interstate coupling.[52] The expressions for the $\mathbf{H}^d$ determined ($x = m$) counterparts are

$$L_0^{I,I,(m)}(\mathbf{R}) = \mathbf{d}^I(\mathbf{R})^\dagger \mathbf{H}^d \mathbf{d}^I(\mathbf{R}), \tag{10}$$

$$L_j^{I,I,(m)}(\mathbf{R}) = \mathbf{d}^I(\mathbf{R})^\dagger (\nabla_j \mathbf{H}^d) \mathbf{d}^I(\mathbf{R}), \tag{11}$$

$$L_j^{I,J,(m)}(\mathbf{R}) = \mathbf{d}^I(\mathbf{R})^\dagger (\nabla_j \mathbf{H}^d) \mathbf{d}^J(\mathbf{R}). \tag{12}$$

To obtain the optimized parameters, we minimize the following performance index:

$$P(\boldsymbol{\theta}) = \frac{1}{2} \sum_{n=1}^{N^{\mathrm{lsq}}} \left[w_n \left(L_n^{(m)} - L_n^{(ab)} \right) \right]^2 + \frac{1}{2} t \boldsymbol{\theta}^\dagger \boldsymbol{\theta}, \tag{13}$$

where N^{lsq} is the number of least squares terms and w_n is the weight for each least squares term. The weights for energy error terms are set to 1, and the weights for energy gradient error and interstate coupling error terms are set as follows:

$$w(L_j^{I,I}) = \sqrt{\frac{\max\left\{ L_0^{I,I,(ab)} \right\}^2}{\max\left\{ L_j^{I,I,(ab)} \right\}^2}}, \tag{14}$$

$$w(L_j^{I,J}) = \sqrt{\frac{\max\left\{ L_0^{I,I,(ab)} \right\}^2}{\max\left\{ L_j^{I,J,(ab)} \right\}^2}}. \tag{15}$$

Including derivative information can help alleviate the tendency of overfitting.[42,43] Still, the regularization term $\frac{1}{2} t \boldsymbol{\theta}^\dagger \boldsymbol{\theta}$ is added to the performance index, where t is a small positive factor (e.g., 10^{-6}). The derivatives of Eqs. (10)–(12) with respect to NN parameters can be analytically evaluated; therefore, the gradient-based optimization algorithms such as the conjugate gradient method and quasi-Newton can be employed to find optimal NN parameters. In this work, the Levenberg-Marquardt algorithm is used to minimize the performance index, which is very numerically robust and can achieve convergence very quickly.[53] To obtain the best results, multiple trainings with different initial parameters are performed (20 trainings in this work), from which the fitting with the smallest performance index is selected as the optimal result.

D. Neural networks with symmetry adaptation

As discussed previously,[19] the 1, 2^1A states of NH_3 carry the A_1' and A_2'' irreducible representations of the CNPI group $\mathbf{G}_{12}$ isomorphic to $S_3 \otimes I$ (direct product of the symmetric group of order 3 and the inversion group). The characters of both irreducible representations are listed in Table I. As both representations are one-dimensional, the two-state $\mathbf{H}^d$ will have three unique symmetry blocks, each carrying a one-dimensional irreducible representation. The two diagonal blocks carry the A_1' irreducible representation, which means that H_{11}^{d} and H_{22}^{d} are invariant with respect to inversion of all nuclei and electrons and permutations of three hydrogen nuclei. On the other hand, the off-diagonal block carries the A_2'' irreducible representation.

The permutation invariant polynomials (PIPs) carry the A_1' irreducible representation of $\mathbf{G}_{12}$ and can be used as inputs of an NN to account for the A_1' symmetry for the two diagonal blocks.[32,46] To obtain a set of PIPs, the monomials are defined as single-valued functions of internuclear distances. For example, Bowman and co-workers used the Morse-like variables in their PIP fitting approach[49,50]

$$p_{ij} = \exp(-a r_{ij}), \tag{16}$$

where a is an adjustable parameter and r_{ij} and p_{ij} represent the bond length between the ith and jth atoms and the corresponding monomials. Other choices of the monomials include inverse ($1/r_{ij}$, used in this work) and logarithmic functions. Based on the monomials, the PIPs can be readily derived as

TABLE I. Characters of A_1' and A_2'' irreducible representations of $S_3 \otimes I$.[a]

	E	(123), (132)	(12), (23), (13)	E*	(123)*, (132)*	(12)*, (23)*, (13)*
A_1'	1	1	1	1	1	1
A_2''	1	1	−1	−1	−1	1

[a]The star symbol * denotes inversion operation.

$$\mathbf{G} = \hat{S}\prod_{i<j}^{N} p_{ij}^{l_{ij}}, \tag{17}$$

where l_{ij} is the order of each monomial and N is the number of atoms. $\hat{S}$ is the symmetrization operator, which consists of the symmetric sum over all possible permutation operations in this group. It is important to note that since internuclear distances are invariant with respect to inversion, so are the resulting PIPs. The more precise name for these polynomials would be permutation inversion invariant polynomials (PIIPs). However, the conventional name PIP will still be used.

It has been demonstrated previously[19] that the dot-cross product function

$$Q^{(3)}_{NHHH} = \frac{(\mathbf{R}_N - \mathbf{R}_{H_1})\cdot(\mathbf{R}_N - \mathbf{R}_{H_2})\times(\mathbf{R}_N - \mathbf{R}_{H_3})}{r_{NH_1}r_{NH_2}r_{NH_3}} \tag{18}$$

and its odd order terms transform as A_2'' irreducible representation. Within the NN framework, one can still take advantage of the PIPs in representing the A_2'' off-diagonal term. This can be done by expressing such a term as a product of $Q^{(3)}_{NHHH}$ (or its odd order terms) and an NN of A_1' symmetry.[46]

The functional form for three blocks are given by

$$H^{\mathrm{d}}_{11} = \mathrm{NN}_1(\mathrm{PIP}), \tag{19}$$

$$H^{\mathrm{d}}_{12} = Q^{(3)}_{NHHH}\mathrm{NN}_2(\mathrm{PIP}) + [Q^{(3)}_{NHHH}]^3\mathrm{NN}_3(\mathrm{PIP}), \tag{20}$$

$$H^{\mathrm{d}}_{22} = \mathrm{NN}_4(\mathrm{PIP}), \tag{21}$$

where $\mathrm{NN}_i(\mathrm{PIP})$ (i = 1–4) are neural network functions, which take PIPs as input. There are two terms for the off-diagonal block H^{d}_{12}, which provide more functional flexibility to achieve better fitting results.

E. Description near conical intersections: Partially diagonalized representation

Orthogonal intersection adapted coordinates[54,55] are used to describe the PESs in the vicinity of a conical intersection seam, which are based on the $\mathbf{g}^{I,J}$ and $\mathbf{h}^{I,J}$ vectors,

$$2g_k^{I,J,(x)}(\mathbf{R}) = L_k^{J,J,(x)}(\mathbf{R}) - L_k^{I,I,(x)}(\mathbf{R}), \tag{22}$$

$$h_k^{I,J,(x)}(\mathbf{R}) = L_k^{I,J,(x)}(\mathbf{R}), \tag{23}$$

where $(x) = m$ or ab, which will be suppressed in the following discussions. The local topography of points on a conical intersection seam is determined by the conical parameters, $s_w^{I,J}$ ($w = x$ or y denotes projection of $\mathbf{s}$ along $\mathbf{g}$ or $\mathbf{h}$ vectors) and $g^{I,J}$, $h^{I,J}$, where

$$2s_k^{I,J}(\mathbf{R}) = L_k^{J,J}(\mathbf{R}) + L_k^{I,I}(\mathbf{R}), \tag{24}$$

$$g^{I,J} = ||\mathbf{g}^{I,J}||, \tag{25}$$

$$h^{I,J} = ||\mathbf{h}^{I,J}||. \tag{26}$$

These parameters are continuous along the conical intersection seam and therefore readily described by $\mathbf{H}^d$, provided

$$\mathbf{g}^{I,J}(\mathbf{R})\cdot\mathbf{h}^{I,J}(\mathbf{R}) = 0. \tag{27}$$

This requirement is satisfied by the proper choice of a one parameter rotation of the degenerate states.[55]

To obtain the derivatives of Eqs. (10)–(12) with respect to NN parameters, the derivatives $(\mathbf{d}^I)^\dagger\frac{\partial}{\partial\theta_k}\mathbf{d}^J \equiv D_k^{I,J}$ are required. Away from the conical intersection seam, evaluation of $D_k^{I,J}$ is readily given by

$$D_k^{I,J} = \left(\mathbf{d}^{I\dagger}\frac{\partial\mathbf{H}^d}{\partial\theta_k}\mathbf{d}^J\right)(E^{a,J,(m)} - E^{a,I,(m)})^{-1}. \tag{28}$$

However, at a conical intersection, Eq. (28) is singular and Eq. (27) is used.[21] In that case, $D_k^{I,J}$ is determined by differentiating Eq. (27), which yields

$$\begin{aligned} D_k^{I,J}\left\{2\left[(h^{IJ})^2 - (g^{IJ})^2\right]\right\} = \mathbf{h}^{I,J}\cdot\Bigg[&\frac{1}{2}\mathbf{d}^{I\dagger}\frac{\partial\nabla\mathbf{H}^d}{\partial\theta_k}\mathbf{d}^I - \frac{1}{2}\mathbf{d}^{J\dagger}\frac{\partial\nabla\mathbf{H}^d}{\partial\theta_k}\mathbf{d}^J \\ &+ \sum_{K\neq I,J} D_k^{K,I}L^{K,I,(m)} - \sum_{K\neq I,J} D_k^{K,J}L^{K,J,(m)}\Bigg] \\ -\mathbf{g}^{I,J}\cdot\Bigg[&\mathbf{d}^{I\dagger}\frac{\partial\nabla\mathbf{H}^d}{\partial\theta_k}\mathbf{d}^J + \sum_{K\neq I,J} D_k^{K,I}L^{K,J,(m)} \\ &+ \sum_{K\neq I,J} D_k^{K,J}L^{K,I,(m)}\Bigg]. \end{aligned} \tag{29}$$

For polyatomic molecules, the dimensionality of the seam space is $N^{\mathrm{int}} - 2$, where N^{int} is the number of internal degrees of freedom. To locate and fit the conical intersection seam in its vicinity can be expensive and inefficient in complex systems, where $N^{\mathrm{int}} - 2$ is large. Furthermore, near a conical intersection, small changes in θ_k would lead to rapid changes in $\mathbf{d}^J$, making the evaluation of Eq. (28) numerically unstable. To this end, a partially diagonalized representation[22] is introduced to facilitate a better description of near degeneracy data points.

The states I, J are defined as quasidegenerate if $|\Delta E^{a,J,I,(ab)}(\mathbf{R})| = |E^{a,J,(ab)}(\mathbf{R}) - E^{a,I,(ab)}(\mathbf{R})|$ is less than a preassigned value (0.05 eV in this work), and are said to form a (quasi) degeneracy group. For a state pair in the same degeneracy group at a point $\mathbf{R}$, Eq. (27) is used to redefine the results of Eq. (3), while any remaining states at $\mathbf{R}$ would satisfy Eq. (3). In a degeneracy group, *in lieu* of using Eq. (28) to evaluate $\mathbf{D}^{I,J}$, Eq. (29) is used. Instead of simply rotating the states as in true degeneracy case, the finite energy difference is taken into account to ensure correct treatment. This provides a partially diagonal representation for which $\mathbf{D}^{I,J}$ are well behaved, facilitating convergence of nonlinear optimization of neural networks. The details of partial diagonalization and $\nabla \mathbf{d}^J$ in the partial diagonalization representation have been reported previously.

F. Neural network for NH_3

The *ab initio* data for NH_3 were used in a previous symmetrized polynomial fitting, and the corresponding results will be denoted as SURFGEN.[56] A total of 4344 geometries giving rise to 85 442 least squares terms are assembled. The NN structure is 30-160-80-4, which has two hidden layers and 18 164 parameters. The transfer function in the first and second layers is a hyperbolic tangent function $f(x) = \tanh(x)$; in the third layer, it is a linear function $f(x) = x$. 30 PIPs are generated for the NN formulation by the Effective Monomial Symmetrization Approach (EMSA) program[57] and are used as the input of this neural network. The 4 output components a_i^3 (i = 1–4) correspond to NN_i (i = 1–4) in Eqs. (19)–(21). One could also use 4 individual NNs to express NN_i; the NN structure used here has a more compact form.

TABLE II. Weights of each energy range in the least squares fitting procedure.

Energy range (cm^{-1})	Weight
0–50 000	1.0
50 000–55 000	0.5
55 000–60 000	0.2
60 000–70 000	0.1
70 000–1 00 000	0.03
1 00 000–1 000 000	0.01
>1 000 000	0.001

III. PERFORMANCE OF THE NN-BASED H^d

A. Energies and comparison with previous work

The inaccessible high energy regions have little influence on dynamics simulations; therefore, in NN fitting, the least squares terms with high energies are given reduced weights, which decrease with increasing energy ranges. Weights used for each energy bracket can be found in Table II. Table III reports the comparison between SURFGEN and NN fits in terms of RMSE (root mean square error) and mean unsigned errors (MUE) for each energy bracket. Considering the energy brackets, the accuracy of each surface is quite satisfactory. Due to the reduced weights for high energy least squares terms, the errors are increasing as the energy ranges become higher. The energy errors of SURFGEN and NN fitting are comparable. It is important to note that, compared to SURFGEN, NN fitting has improved the errors for most energy brackets.

TABLE III. Root mean square error (RMSE) (cm^{-1}) and mean unsigned error (MUE) (cm^{-1}) of SURFGEN and NN fitting for various energy brackets of ground and excited states ($E^{a,J,(m)} = E_J$, J = 1, 2). Energies in cm^{-1} relative to $NH_3(\tilde{X})$ minimum.

Energy range	Number of data points	MUE (SURFGEN)	RMSE (SURFGEN)	MUE (NN)	RMSE (NN)
		E_1			
0–10 000	186	13.96	17.28	7.07	9.83
0–20 000	511	16.00	20.41	12.30	17.82
0–30 000	1058	18.34	24.18	17.50	24.84
0–40 000	1961	22.20	29.76	19.76	27.47
0–50 000	3138	27.78	38.73	20.02	27.17
0–60 000	3946	34.11	50.80	25.65	39.15
0–70 000	4124	40.27	76.09	32.84	79.98
0–80 000	4231	54.11	159.39	45.65	149.44
0–90 000	4295	64.53	218.32	53.47	186.12
		E_2			
0–50 000	257	19.18	25.75	11.38	18.30
0–60 000	1335	57.56	91.55	49.60	81.04
0–70 000	2260	116.62	199.21	117.54	208.09
0–80 000	3243	247.15	451.11	240.85	433.54
0–90 000	3968	324.23	559.61	313.42	532.49

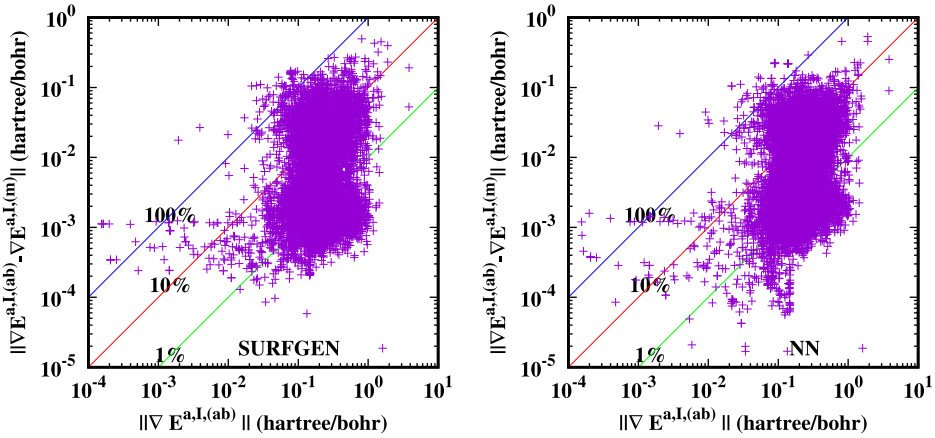

FIG. 2. Error in $\mathbf{H}^d$ determined energy gradients for states I = 1, 2. $||\nabla E^{a,I,(m)}(\mathbf{R}) - \nabla E^{a,I,(ab)}(\mathbf{R})||$ plotted on the ordinate against the *ab initio* energy gradients $||\nabla E^{a,I,(ab)}(\mathbf{R})||$ on the abscissa. A log-log scale is used. 45° lines represent 100%, 10%, and 1% error, respectively.

B. Energy gradients and derivative couplings

The analysis of the quality of the NN-based Hamiltonian in terms of energy gradient errors is shown in Fig 2, which presents plots of the norm of the error of the energy gradients against the norm of the corresponding *ab initio* quantity for both SURFGEN and NN fitting. Since the scales are logarithmic, the 45° lines represent 100% (blue line), 10% (red line), and 1% (green line) error, respectively. As can be seen in Fig. 2, energy gradients of the majority of the points are below the 10% error line for both SURFGEN and NN fitting. In general, NN fitting has also improved the errors of gradients compared to SURFGEN. The average error for gradients in NN fitting is 16.8%, while it is 25.6% in SURFGEN.

Figure 3 presents a path connecting the minimum energy point on the 2^1A potential energy surface to the saddle point on that surface, the saddle point to the minimum energy point on the 1, 2 ^1A state conical intersection seam, and the seam point to the product channel. This path is essential for the description of the nonadiabatic photodissociation process, given by reaction (1). Along this path, apart from the region near the conical intersection seam, the reproduction of derivative couplings in NN fitting is quite satisfactory. Least squares fitting cannot exactly reproduce the location of conical intersections, and it is the interstate couplings (**h** vector) that are fit, rather than derivative couplings; still, the *ab initio* derivative coupling is well reproduced by the fit although its peak is slightly shifted. Figure 4 presents the norms of the **g** and **h** vectors along the same dissociative path near the minimum energy crossing (MEX). As can be seen, the ||**g**|| is well reproduced by NN fitting. However, there is a small but noticeable difference at small r_{NH} distances for ||**h**||, which is probably due to the nonremovable part in derivative couplings. In general, both the **g** and **h** vectors are well-reproduced near the conical intersection (r_{NH} = 3.72 bohrs)

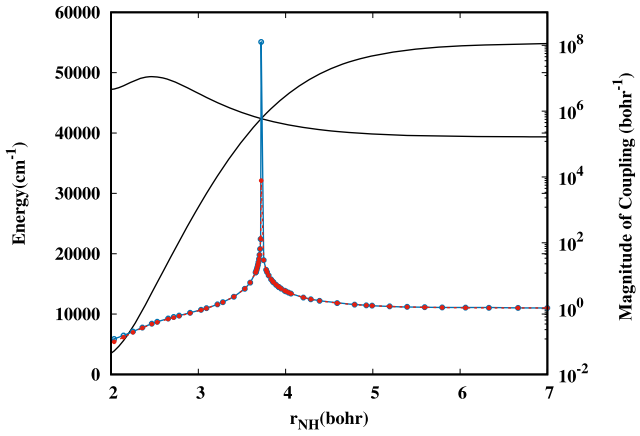

FIG. 3. Plot of energies $E^{a,I,(ab)}$, $E^{a,J,(ab)}$ (black solid lines), derivative coupling norms $||\mathbf{f}^{a,I,J,(ab)}||$ (blue solid line with open circles) and $||\mathbf{f}^{a,I,J,(m)}||$ from NN fitting (red dashed line with filled circles) as a function of r_{NH} along a dissociative path which is the union of points connecting the $\tilde{A}$ state minimum to the $\tilde{A}$ state saddle point, the $\tilde{A}$ state saddle point to the minimum energy crossing (MEX) points, and the MEX point toward the asymptote.

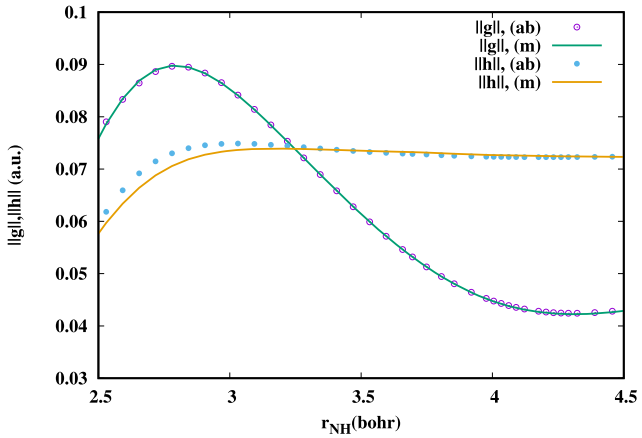

FIG. 4. Plot of ||**g**|| and ||**h**|| as functions of r_{NH} near minimum energy crossing (MEX) along the same dissociative path as in Fig 3. *Ab initio* determined results are marked in symbols, and the corresponding NN fitting results are shown in lines.

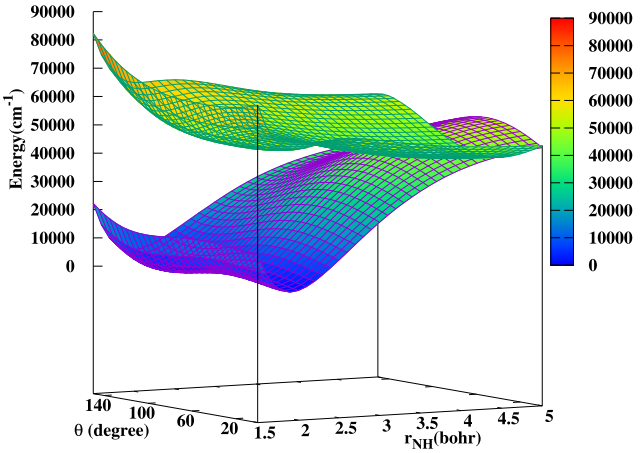

FIG. 5. 3D plot of H^d_{11} and H^d_{22} functions of r_{NH} and out-of-plane angle θ. The reference geometry is the minimum of the $\tilde{X}$ state. θ is the angle between the dissociative N–H bond and the C_3 axis.

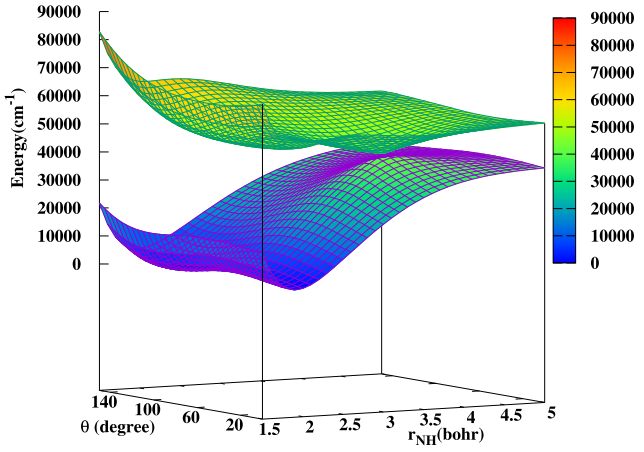

FIG. 7. 3D plot of $\mathbf{H}^d$ determined adiabatic energies as a function of r_{NH} and out-of-plane angle θ. The reference geometry is the minimum of the $\tilde{X}$ state. θ is the angle between the dissociative N-H bond and the C_3 axis.

so that the local topography of the conical intersection seam is well described.

Figures 5–7 present 3D plots of the NN-based quasidiabatic $H^d_{1,1}$, $H^d_{2,2}$, $H^d_{1,2}$ and the $\mathbf{H}^d$ determined adiabatic energies, which give a more global picture of the quasidiabatic representation. The conical intersection between $1,2^1A$ states is clearly reproduced in Fig. 7. Although the scale is relatively large for $H^d_{1,1}$, $H^d_{2,2}$ and adiabatic energies, no figure shows any sign of oscillations and the smoothness is evident. The analytical form of a feed-forward neural network allows massively parallel evaluation of NN-based $\mathbf{H}^d$. The evaluations of energies, energy gradients, and derivative couplings are basically matrix manipulations. These procedures are readily implemented in LAPACK[58] subroutines, which are fully optimized and freely available on most architectures. It took only 4 s of central processing unit (CPU) time to evaluate 10 000 geometries on a 24-core machine.

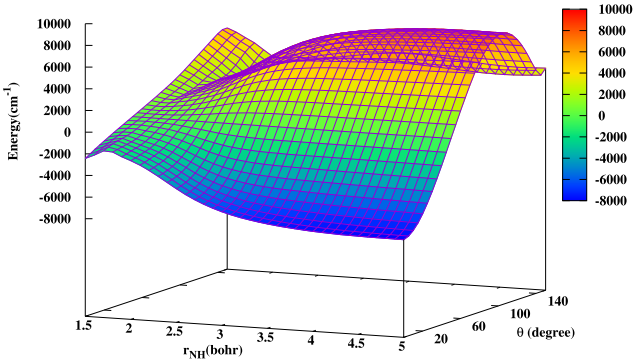

FIG. 6. 3D plot of H^d_{12} as a function of r_{NH} and out-of-plane angle θ. The reference geometry is the minimum of the $\tilde{X}$ state. θ is the angle between the dissociative N–H bond and the C_3 axis.

IV. SUMMARY

A previously reported fitting procedure to generate NN-based quasidiabatic Hamiltonians has been generalized. Symmetry-adapted permutation invariant polynomials are used to account for complete nuclear permutation inversion (CNPI) symmetry. A partially diagonalized representation is used to facilitate a better description of near degeneracy points. The new fitting procedure was tested using *ab initio* data originally used to describe the ammonia photodissociation process, reaction (1). The errors in energies and energy gradients are improved compared to a previous polynomial fitting. The behavior of the derivative couplings near a seam of the conical intersection is very well reproduced. The well-reproduced **g** and **h** vectors guarantee the reproduction of the local topology of the conical intersection seam.

In summary, the NN-based quasidiabatic Hamiltonian with symmetry adaptation that can reproduce the local topography of a seam of conical intersections is constructed by a least squares fitting method. The flexible functional form of NNs allows accurate reproduction of *ab initio* data including energies, energy gradients, and the derivative couplings.

Ammonia is a relatively simple example, which only needs a limited set of geometries to describe the relevant PESs. Thus, a relatively small number of least square terms are required, which can be handled by the Levenberg-Marquardt algorithm. However, as the system under study becomes larger and more complicated, the number of least square terms will be significantly larger, requiring more complex NN structure to fit. In that case, the Levenberg-Marquardt algorithm will no longer be suitable because it scales cubically with respect to the number of nonlinear parameters and needs to store the whole Jacobian matrix. Therefore, future work will focus on the application of neural networks to larger and more complicated systems using nontraditional training approaches.

SUPPLEMENTARY MATERIAL

The NN determined $\mathbf{H}^d$ (PEM) reported in this work is provided in the supplementary material.

ACKNOWLEDGMENTS

Y.G., H.G., and D.R.Y. acknowledge support of the Department of Energy, Grant No. DE-SC0015997. Computer time was provided by NERSC.

There are no conflicts of interest to declare.

REFERENCES

[1] F. Gatti, *Molecular Quantum Dynamics: From Theory to Applications* (Springer, Berlin, 2014).
[2] H. Guo and D. R. Yarkony, Phys. Chem. Chem. Phys. **18**, 26335 (2016).
[3] M. Baer, Chem. Phys. **15**, 49 (1976).
[4] C. A. Mead and D. G. Truhlar, J. Chem. Phys. **77**, 6090 (1982).
[5] M. Baer, Phys. Rep. **358**, 75 (2002).
[6] J. E. Subotnik, S. Yeganeh, R. J. Cave, and M. A. Ratner, J. Chem. Phys. **129**, 244101 (2008).
[7] C. E. Hoyer, K. Parker, L. Gagliardi, and D. G. Truhlar, J. Chem. Phys. **144**, 194101 (2016).
[8] H. Nakamura and D. G. Truhlar, J. Chem. Phys. **117**, 5576 (2002).
[9] H. Nakamura and D. G. Truhlar, J. Chem. Phys. **118**, 6816 (2003).
[10] T. Lenzen and U. Manthe, J. Chem. Phys. **147**, 084105 (2017).
[11] D. M. G. Williams and W. Eisfeld, J. Chem. Phys. **149**, 204106 (2018).
[12] H. Köppel, "Diabatic representation: Methods for the construction of diabatic electronic states," in *Conical Intersections* (World Scientific, Singapore, 2011), pp. 175–204.
[13] R. G. Sadygov and D. R. Yarkony, J. Chem. Phys. **109**, 20 (1998).
[14] R. Abrol and A. Kuppermann, J. Chem. Phys. **116**, 1035 (2002).
[15] D. Yuan, Y. Guan, W. Chen, H. Zhao, S. Yu, C. Luo, Y. Tan, T. Xie, X. Wang, Z. Sun, D. H. Zhang, and X. Yang, Science **362**, 1289 (2018).
[16] C. R. Evenhuis and M. A. Collins, J. Chem. Phys. **121**, 2515 (2004).
[17] O. Godsi, C. R. Evenhuis, and M. A. Collins, J. Chem. Phys. **125**, 104105 (2006).
[18] C. Evenhuis and T. J. Martínez, J. Chem. Phys. **135**, 224110 (2011).
[19] X. Zhu and D. R. Yarkony, J. Chem. Phys. **132**, 104101 (2010).
[20] X. Zhu and D. R. Yarkony, J. Chem. Phys. **136**, 174110 (2012).
[21] X. Zhu and D. R. Yarkony, J. Chem. Phys. **137**, 22A511 (2012).
[22] X. Zhu and D. R. Yarkony, J. Chem. Phys. **140**, 024112 (2014).
[23] J. Ma, X. Zhu, H. Guo, and D. R. Yarkony, J. Chem. Phys. **137**, 22A541 (2012).
[24] J. Ma, C. Xie, X. Zhu, D. R. Yarkony, D. Xie, and H. Guo, J. Phys. Chem. A **118**, 11926 (2014).
[25] C. Xie, J. Ma, X. Zhu, D. H. Zhang, D. R. Yarkony, D. Xie, and H. Guo, J. Phys. Chem. Lett. **5**, 1055 (2014).
[26] C. Xie, X. Zhu, J. Ma, D. R. Yarkony, D. Xie, and H. Guo, J. Chem. Phys. **142**, 091101 (2015).
[27] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, J. Chem. Phys. **103**, 4129 (1995).
[28] D. F. R. Brown, M. N. Gibbs, and D. C. Clary, J. Chem. Phys. **105**, 7597 (1996).
[29] C. M. Handley and P. L. A. Popelier, J. Phys. Chem. A **114**, 3371 (2010).
[30] J. Behler, Phys. Chem. Chem. Phys. **13**, 17930 (2011).
[31] J. Chen, X. Xu, X. Xu, and D. H. Zhang, J. Chem. Phys. **138**, 221104 (2013).
[32] B. Jiang and H. Guo, J. Chem. Phys. **139**, 054112 (2013).
[33] J. Li, B. Jiang, and H. Guo, J. Chem. Phys. **139**, 204103 (2013).
[34] B. Jiang and H. Guo, J. Chem. Phys. **141**, 034109 (2014).
[35] S. Manzhos, R. Dawes, and T. Carrington, Int. J. Quantum Chem. **115**, 1012 (2015).
[36] B. Jiang and H. Guo, Phys. Rev. Lett. **114**, 166101 (2015).
[37] X. Shen, J. Chen, Z. Zhang, K. Shao, and D. H. Zhang, J. Chem. Phys. **143**, 144701 (2015).
[38] T. Liu, Z. Zhang, B. Fu, X. Yang, and D. H. Zhang, Chem. Sci. **7**, 1840 (2016).
[39] K. Shao, J. Chen, Z. Zhao, and D. H. Zhang, J. Chem. Phys. **145**, 071101 (2016).
[40] B. Kolb, X. Luo, X. Zhou, B. Jiang, and H. Guo, J. Phys. Chem. Lett. **8**, 666 (2017).
[41] K. Shakouri, J. Behler, J. Meyer, and G.-J. Kroes, J. Phys. Chem. Lett. **8**, 2131 (2017).
[42] A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum, and R. Komanduri, J. Chem. Phys. **130**, 134101 (2009).
[43] H. T. Nguyen-Truong and H. M. Le, Chem. Phys. Lett. **629**, 40 (2015).
[44] K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).
[45] Y. Guan, B. Fu, and D. H. Zhang, J. Chem. Phys. **147**, 224307 (2017).
[46] C. Xie, X. Zhu, D. R. Yarkony, and H. Guo, J. Chem. Phys. **149**, 144107 (2018).
[47] Y. Guan, D. H. Zhang, H. Guo, and D. R. Yarkony, Phys. Chem. Chem. Phys. (published online, 2019).
[48] P. R. Bunker and P. Jensen, *Molecular Symmetry and Spectroscopy* (NRC Research Press, 2006).
[49] B. J. Braams and J. M. Bowman, Int. Rev. Phys. Chem. **28**, 577 (2009).
[50] J. M. Bowman, G. Czakó, and B. Fu, Phys. Chem. Chem. Phys. **13**, 8094 (2011).
[51] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[52] D. R. Yarkony, J. Chem. Phys. **114**, 2601 (2001).
[53] M. T. Hagan and M. B. Menhaj, IEEE Trans. Neural Networks **5**, 989 (1994).
[54] G. J. Atchity, S. S. Xantheas, and K. Ruedenberg, J. Chem. Phys. **95**, 1862 (1991).
[55] D. R. Yarkony, J. Chem. Phys. **112**, 2111 (2000).
[56] See https://github.com/virtualzx-nad/NH3-X-A-Coupled-PES for the usage of the SURFGEN PEM.
[57] Z. Xie and J. M. Bowman, J. Chem. Theory Comput. **6**, 26 (2010).
[58] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).