Determination of Resonant State Parameters from Ab Initio Data by Artificial Neural Network and Statistical Pade Approximation*)

Lukáš PICHL and Jiří HORÁČEK[1)]

International Christian University, 3-10-2 Osawa, Mitaka, Tokyo 181-8585, Japan

[1)]*Institute of Theoretical Physics, Charles University Prague, V Holešovičkách 2, 180 00 Praha 8, Czech Republic*

(Received 27 October 2014 / Accepted 8 March 2015)

Resonant states in electron molecule collisions mediate a variety of energy transfer processes in the plasma edge region, such as vibrational excitation, dissociative recombination, dissociative attachment, associative detachment etc. Here we show that the resonance parameters, in general difficult to obtain, can be computed from standard bound-state ab initio data by means of analytical continuation in the coupling constant. The procedure uses artificial neural network and statistical Pade approximation to extrapolate from the bound-state region to that of the resonant state by varying the strength of the attractive potential term added. We present benchmark data for the ethylene molecule and demonstrate a reasonable stability of the results over the quantum chemical basis sets employed.

Keywords: fusion edge plasma, electron-molecule collision, resonance energy, resonance width, artificial neural network, statistical Pade approximation

DOI: 10.1585/pfr.10.3406049

1. Introduction

Electron-molecule collisions belong among the important processes taking place in fusion edge plasma through which the energy is distributed in the divertor region of fusion reactors. We are interested in the processes of dissociative attachment (electron-impact induced dissociation of a target molecule accompanied by electron attachment) and the reverse process of associative detachment (formation of a bond among a target colliding with an anion accompanied by electron detachment into the continuum). These processes are mediated by a resonant state. This resonant state is the key mediator of the dynamical processes such as the associative detachment (AD) and dissociative attachment (DA), which involve anions temporarily formed during the collision process. The parameters of the resonant state are the essential input parameters to the dynamical models that allow the subsequent determination of the AD and DA cross sections [1]. One approach to determination of the location and the lifetime of the resonance is an ab-initio calculation using an extra attractive potential, coupling term,

$$V(r) \to V(r) + \lambda U(r), \tag{1}$$

which renders the resonant state to the bound-state regime.

The strength of the extra attractive term, i.e. the coupling constant λ, is the key parameter for the method of analytical continuation that determines the resonance parameters. In practice, the coupling term is usually an extra charge on a Coulomb center in the molecule. The resonant state is extrapolated from the series of the bound-state values in the limit of $\lambda = 0$ [2]. Taking $E = \kappa^2$, we obtain a series of bound-state values from the ab initio calculation, results of which are sampling the function $\kappa(\lambda)$ for $\lambda \in \{\lambda_i\}$, $i = 1 \dots N$. This is an increasing function, and thus its inverse, $\lambda(\kappa)$ exists. Both functions, $\lambda(\kappa)$ and $\kappa(\lambda)$, will be utilized in the method of analytical continuation.

2. Theoretical Method

In order to explain the method of the analytical continuation in the coupling constant λ, we now introduce the basic tools - statistical Pade approximation [2] and the artificial neural network (ANN) [3]. Pade approximation to a function $f(x)$ is a rational polynomial function,

$$f(x) \simeq \frac{P_n(x)}{Q_m(x)}, \quad P_n(x) = \sum_{i=0}^{n} a_i x^i, \tag{2}$$

$$Q_m(x) = 1 + \sum_{i=1}^{m} b_i x^i.$$

In order to obtain the coefficients a_i and b_i of the polynomials, one has to solve the nonlinear least square problem,

$$\sum_{i=1}^{N} \left| f_i - \frac{P_n(x_i)}{Q_m(x_i)} \right|^2 \to \min. \tag{3}$$

In order to initialize the procedure of the iterative solution, we can first minimize

$$\sum_{i=1}^{N} |f_i Q_m(x_i) - P_n(x_i)|^2 ,$$

author's e-mail: lukas@icu.ac.jp

*) This article is based on the presentation at the 24th International Toki Conference (ITC24).

which in turn results in a linear system of equations for $\mathbf{z} = \{a_0, a_1, \ldots a_n, b_1, \ldots, b_m\}$. We denote the matrix of the coefficients as M. This over-determined linear system,

$$\sum_{j=1}^{s} m_{i,j} z_j = y_i, \quad i = 1 \ldots N, \quad N > s \tag{4}$$

can be solved by the singular value decomposition (SVD) of the coefficient matrix M,

$$M = U \Sigma V^*, \tag{5}$$

where U is an $N \times N$ unitary matrix, Σ an $N \times s$ diagonal matrix of rectangular shape (with non-negative values on the diagonal), V an $s \times s$ unitary matrix, $s = n + m + 1$, and the asterisk denotes its conjugate transpose. By means of the matrix pseudoinverse based on Eq. (5), the unique solution of Eq. (4) is routinely obtained, subject to

$$\sum_{i=1}^{N} \left(\sum_{j=1}^{s} M_{ij} z_j - y_i \right)^2 \to \min, \tag{6}$$

which is the linear least square form for the residues. For the nonlinear solution, this initial guess is further iterated. The solution is obtained from the weighted least square problem, where the weights ϵ_i for each residue are given as $1/\epsilon_i^{(t)} = Q_m^{(t-1)}(x_i)$ (t denotes the index of iteration). To stabilize the rate of convergence, we also modify the iteration scheme by a standard relaxation technique, averaging the last two results to obtain the new iterand.

2.1 Analytical continuation in the coupling constant

In order to compute the complex value of $\kappa(0)$, and thus obtain the resonance parameters, there are two approaches possible. We first perform the direct statistical Pade approximation for the function $\lambda(\kappa)$,

$$\lambda(\kappa) \simeq \frac{P_n(\kappa)}{Q_m(\kappa)}, \tag{7}$$

and solve for the roots of the polynomial in the numerator, i.e.

$$P_n(\kappa_c) = 0. \tag{8}$$

In this approach, there are n complex roots, and the stable solution is selected by tracking the results for different orders of Pade approximation. Since the numerical problem of the multipoint, or statistical, Pade approximation is ill-conditioned, some care must be taken in the selection of points, suitable order of the Pade approximation, and overall numerical precision of the calculation. The complex energy of the resonance is obtained as $E_c = \kappa_c^2$.

Next, we proceed to the inderect approach, making better use of the analytical structure of the problem. First, we determine the value of $\lambda_0 \equiv \lambda(0)$ for $\kappa = 0$, i.e. the threshold value for which a bound state still exists. It is known that κ is analytical in terms of $\sqrt{\lambda - \lambda_0}$ [4]. For the purpose of analytical continuation, we can now adopt e.g. Pade approximation anew as

$$\kappa\left(\sqrt{\lambda - \lambda_0}\right) \simeq \frac{P'_n(\sqrt{\lambda - \lambda_0})}{Q'_m(\sqrt{\lambda - \lambda_0})} \tag{9}$$

(here P' and Q' denotes polynomials different from those of the direct method). Alhough being sensitive to the precision level of λ_0, this Pade approximation is otherwise quite stable, and the resonance parameter κ_c is obtained directly as a single value from the straightforward substitution,

$$\kappa_c = \frac{P'_n(\sqrt{-\lambda_0})}{Q'_m(\sqrt{-\lambda_0})}. \tag{10}$$

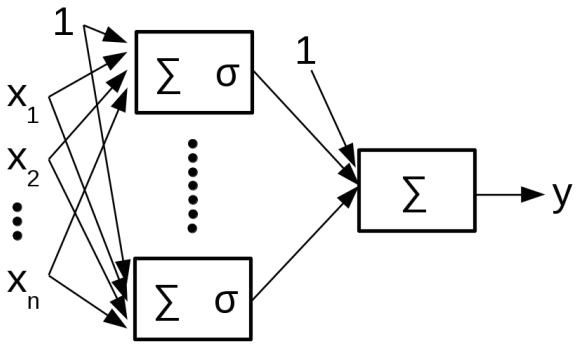

Fig. 1 Artificial neural network in the feed-forward configuration with a single hidden layer and scalar output y. Each synaptic connection denoted by an arrow has an associated weight parameter. Input aggregation is denoted with the symbol Σ whereas σ stands for the nonlinear activation function. The bias value for each neuron is represented by the constant input 1.

2.2 Artificial neural network

In order to obtain λ_0 we could have used the value of a_0 from the direct method. It is, however, preferrable to extrapolate this value by an independent method. We opt for the feed-forward artificial neural network with a single-hidden layer [5] to learn the function $\lambda(\kappa)$ and to determine λ_0. A single hidden layer with a sufficient number of hidden neurons suffices as a universal approximator of arbitrary analytical function [3]. The output of the neural network is known to represent the statistical mean value in regard to the noise hidden in the data [5]. Last but not least, neural networks are known from practice to represent rather well data not covered in the training samples, provided sufficient care is taken to avoid data overfitting. The ANN architecture adopted in this paper is schematically depicted in Fig. 1. Correspondingly, the propagation of input data through the neural network is described by the formula

$$\hat{y}(\boldsymbol{\theta}) = \sigma^2 \left[\sum_{i=1}^{nh} w_i^2 \sigma^1 \left(\sum_{j=1}^{n} w_{i,j}^1 x_j + b_i^1 \right) + b^2 \right], \tag{11}$$

where $\hat{y}$ is the output value computed by the network, $\boldsymbol{\theta}$ denotes collectively all the parameters, i.e. the biases of the hidden neurons b_i^1, $i = 1 \ldots nh$, the bias of the output neuron b^2, weights among the input values x_j, $j = 1 \ldots n$ and the hidden neurons, $w_{i,j}^1$, and the weights among the hidden and output neurons, w_i^2. σ^i stands for the neuron activation function in layer i. Notice that the number of the input values for the network is $n = 1$. The input value x_1 corresponds to κ whereas the output value y represents λ. Computational flow of the neural network is shown in Fig. 1. The inputs are weighted, aggregated, subjected to the activation function of hidden neurons, and forwarded to the output layer. The activation functions of the hidden neurons are nonlinear, $\sigma_i^1(z) = \tanh(z)$, whereas the output activation function is commonly an identity, $\sigma^2(z) = z$ [5]. The first selection promotes the speed of convergence of the neural network whereas the second one is suitable for the unscaled output [5]. It is convenient to denote the aggregate input of the hidden neuron i as s_i^1, and the aggregate input of the output neuron as s^2. We also absorb the bias term by setting an extra input, $x_0 = 1$, and interpret the bias term as the weight coefficient (index zero) for this constant input. Then we have for the output of the neurons in the hidden layer,

$$h_i = \sigma^1\left(\sum_{j=0}^{n} w_{i,j}^1 x_j\right), \quad i = 1 \ldots nh, \tag{12}$$

and for the output of the neural network,

$$\hat{y} = \sigma^2\left(\sum_{i=0}^{nh} w_i^2 h_i\right). \tag{13}$$

The maximum number of hidden neurons is set as $nh = 25$ in the calculation. The set of parameters $\boldsymbol{\theta}$ is determined from non-linear optimization of the objective function using the least square method for

$$J^{(t)} = \frac{1}{2}\left[\hat{y}^{(t)} - y_d^{(t)}\right]^2, \tag{14}$$

where t is an index of the training instance $\{\mathbf{x}^{(t)}, y_d^{(t)}\}$, and $\hat{y}^{(t)}$ is the corresponding output of the neural network for the input $\{\mathbf{x}^{(t)}\}$. The value $y_d^{(t)}$ represents the prescribed output in the training instance t. We now summarize the back-propagation algorithm that determines the weight update from the residual error (Eq. (14) as described in [6]). Since the function represented by the neural network is analytical, we obtain by the chain rule for the derivatives the expression for the gradient of the objective function with respect to the generalized weights. First for the set of weights connecting to the output layer,

$$\frac{\partial J}{\partial w_i^2} = -(y_d - \hat{y})\sigma'^2(s^2)h_i \equiv -\delta^2 h_i, \tag{15}$$

then for the set of weights connecting to the hidden layer,

$$\frac{\partial J}{\partial w_{i,j}^1} = -\delta^2 w_i^2 \sigma'^1(s_i^1)x_j. \tag{16}$$

This can be finalized in the same form as above,

$$\frac{\partial J}{\partial w_{i,j}^1} = -\delta_i^1 x_j, \quad \delta_i^1 = \sigma'^1(s_i^1)\delta^2 w_i^2. \tag{17}$$

Here σ'^i denotes the derivative of the activation function in layer i with respect to the argument. The above equations, Eq. (15) and (17), provide the gradient of all the weights in the neural network, the essential component of the back-propagation algorithm. The gradient vector can then be used directly in the steepest descent method or applied more efficiently using the conjugate gradient method for the nonlinear optimization problem of the ANN architecture parameters. We apply the sequential learning method, in which for every training instance, the input data propagate forward through the neural network, the output value is computed, the difference between the desired and obtained output determined, and the value of this residue then propagates backward through the network updating all of its parameters.

3. Results and Discussions

Having outlined the theoretical method, we now benchmark the above algorithm computationally and present the results for ethylene molecule. The resonant state is the $^2B_{2g}$ shape resonance of the C_2H_4 molecule. This is the lowest located resonant state that plays the crucial role for the processes such as AD and DA. Since the ethylene molecule belongs among hydrocarbons that can be formed in the divertor region of tokamak devices, this system is of interest for the plasma fusion community. The stability of the resonance results obtained by the above method is investigated with respect to the quantum- chemical basis sets employed, namely the cc-pvDZ, cc-pvTZ, cc-pvQZ, cc-pv5Z and cc-pv6Z. These are the Dunning's correlation consistent basis sets [7] (and references therein) used in the calculations of electron affinities by the molpro software package [8]. First, we show the curve of $E(\lambda)$ for the largest basis in Fig. 2. The differences of input data with respect to the smaller basis sets employed is shown in Fig. 3. It can be seen from Fig. 3 that pronounced differences arise when λ is small, since more and more basis functions are necessary to capture the delocalized character of the resonant state.

Table 1 shows the results for the direct method. The resonance location varies less than the resonance width. All data were obtained with the Pade(2,2) approximation.

Table 2 shows the values of λ_0 as extrapolated by the artificial neural network, and the overal results for the energy and width of the resonance. With the exception of the resonance width for the cc-pvQz basis, the results are in good accord and demonstrate the good applicability of the indirect method for ethylene molecule, and of the analytical continuation in the coupling constant in general.

Since the location of the resonance varies at the third significant digit, with the exception of the cc-pvDz ba-

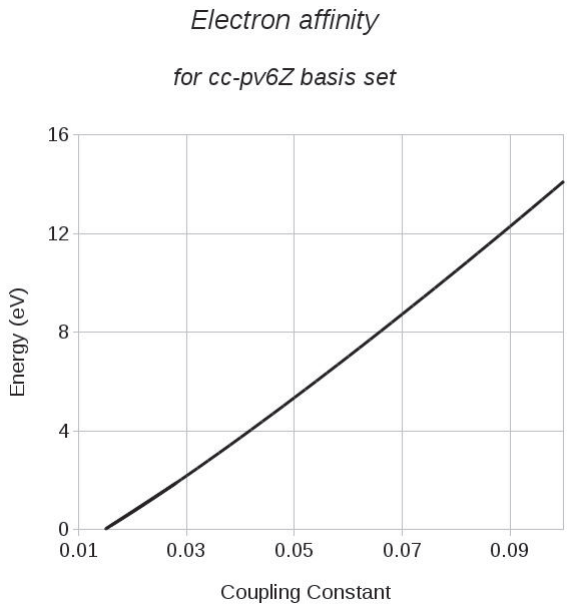

Fig. 2 Electron affinity as a function of coupling constant λ for the cc-pv6z basis.

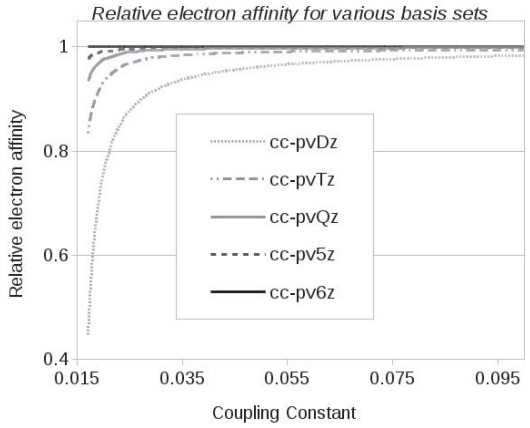

Fig. 3 Electron affinity relative to the cc-pv6z basis value as a function of coupling constant λ.

sis, and the resonance width variation does not exceed 10%, with the exception of cc-pvQz basis, these results are considered quite satisfactory. All values in Table 2 have been obtained with the Pade(n,m) approximation taking $n = 2$ and $m = 2$. In fact, for the cc-pvDz basis, we can obtain a more consistent value $E_{res} = 1.8997$ eV and $\Gamma = 0.39113$ eV within the Pade(4,4) approximation. Similarly, for cc-pvQz, $\Gamma = 0.39725$ eV within the order of Pade(3,4). These diffences may be attributed in part to the noise in the data, and in part to the existence of Froissart doublets, in which case the roots of P' and Q' are located very close to each other, thus presenting a sharp step between zero and a pole of the Pade approximant [9]. We conclude here that the accuracy of the result data within a fixed order of Pade approximation suffices for their application in the nonlocal resonance model of Čížek, Horáček and Domcke [1], and may be used in the boomerang model describing vibrationally inelastic cross sections [10]. This is because the uncertainty in the input ab initio parameters for most previous works has been comparable with that of the present results, which typically translates to the uncertainty in scattering phase shifts of up to 10%. The present work is not only of practical significance for the particular molecular system studied, but also generally relevant to the resonance parameter determination in a variety of electron-molecule collision processes, and thus believed to be of interest for the plasma fusion community.

Table 1 Resonance parameters for the direct method.

Basis set	Energy (eV)	Width (eV)
cc-pvDz	1.957	0.365
cc-pvTz	1.881	0.371
cc-pvQz	1.864	0.410
cc-pv5z	1.844	0.435
cc-pv6z	1.820	0.436

Table 2 Resonance parameters for the indirect method.

Basis set	λ_0	Energy (eV)	Width (eV)
cc-pvDz	0.016133	1.9434	0.39875
cc-pvTz	0.015148	1.8795	0.36737
cc-pvQz	0.014911	1.8167	0.47298
cc-pv5z	0.014715	1.8703	0.38795
cc-pv6z	0.014677	1.8734	0.35461

Acknowledgements

This work was supported by the Grant GACR P203/12/0665. We would like to thank Dr. Ivana Paidarová for providing us the ab initio data on ethylene molecule.

[1] M. Čížek, J. Horáček and W Domcke, J. Phys. B **31**, 2571 (1998).
[2] J. Horáček, P. Mach and J. Urban, Phys. Rev. A **82**, 032713 (2010).
[3] K. Hornik, Neural Neworks **4**, 251 (1991).
[4] V.I. Kukulin, V.M. Krasnopolsky and J. Horáček, *Theory of Resonances: Principles and Applications* (Kluwer Academic Publishers: Dordrecht, The Netherlands, 1988).
[5] W.W. Hsieh, *Machine Learning Methods in Environmental Sciences: Neural Networks and Kernels* (Oxford University Press, Oxford, United Kingdom, 2009).
[6] D.E. Rumelhart, G.E. Hinton and R.J. Williams, Nature **323**, 533 (1986).
[7] A.K. Wilson, T. van Mourik and T.H. Dunning Jr., J. Mol. Struct. (Theochem) **388**, 339 (1996).
[8] H.J. Werner, P.J. Knowles, G. Knizia, F.R. Manby, M. Schutz *et al.*, *MOLPRO: A Package of Ab Initio Programs*, version 2010.1, http://www.molpro.net
[9] J. Gilewicz and Y. Kryakin, J. Compt. Appl. Math. **153**, 235 (2003).
[10] J. Horáček, M. Čížek, K. Houfek, P. Kolorenč and W. Domcke, Phys. Rev. A **73**, 022701 (2006).