Dual vibration configuration interaction (DVCI). A novel factorisation of molecular Hamiltonian for high performance infrared spectrum computation.

Romain Garnier[a,*]

[a] *Queen's University of Belfast*

Abstract

Here is presented an original program based on molecular Schrödinger equations. It is dedicated to target specific states of infrared vibrational spectrum in a very precise way with a minimal usage of memory. An eigensolver combined with a new probing technique accumulate information along the iterations so that desired eigenpairs rapidly tend towards the variational limit. Basis set is augmented from the maximal components of residual vectors that usually require the construction of a big matrix block that here is bypassed with a new factorisation of the Hamiltonian. The latest borrows the mathematical concept of duality and the second quantization formalism of quantum theory.

Keywords: Vibration Configuration Interaction, Infrared spectrum, Iterative eigensolver, Residual error minimization, Duality, Second quantization.

PROGRAM SUMMARY

Program Title: Dual vibration configuration interaction (DVCI)

Licensing provisions: GNU General Public License 3.

Programming languages: C/C++/Fortran.

Supplementary materials:

1. *The sources of the code grouped in folder DualVCI.zip also available at* `https://github.com/4Rom1/DualVCI`
2. *The input files of examples treated in section 6.*

Nature of problem: High computational cost in vibration configuration interaction methods [1, 2], coming from the necessity to solve a large eigenvalue problem to acquire a good precision. The dimension of the matrix exponentially increases with the size of the studied molecule.

* *E-mail address:* rom1{*dot*}garnier{*at*}yahoo{*dot*}fr

Solution method: The A_k decomposition [3] completed by a meaningful error evaluation namely the residue $||HX - EX||$ minimised along iterations for specific targets given in input. The approximation space is generated in the same time as the residual vectors computed on the fly thanks to an adapted choice of excitations shaped on the Hamiltonian operator.

1. Introduction

Nowadays, many devices are able to supply high quality spectrum measurements. However the interpretation of the samples remains a difficult task because the numerical accuracy that is possible to obtain is very limited for medium to large molecules ($>$ 6 atoms). In a typical resolution of vibrational Schrödinger equations in the Born-Oppenheimer frame, one can question the correctness of the model from two principal angles. First, the quality of the results will be affected by the one of the Potential Energy Surface (PES). This last point aside, there remains the validity of the numerical solutions when the PES is provided. That is the focus of the proposed method. A prior analysis of the potential energy supplies the harmonic states that are not able to correctly describe complex combinations of translational and rotational motions of the nucleus. Although still accurately limited, the VSCF method [4, 5] alone allows a better representation. With the idea that these first estimations can be combined to give a more authentic description, the cheapest technique remains the perturbation theory [6, 7, 8] that is known to struggle with strong resonances conventionally encountered in molecular spectroscopy [9, 10]. Vibration configuration interaction (VCI) method [1, 2, 11] permits a better precision with a much higher computational expense coming from the size of the variational space that is tenfold with the number of atoms of the molecule. To avoid this bottleneck, the A_k decomposition [3] initially designed for electronic structure calculation and firstly introduced in the VCI context with the Vibrational Multi Reference Configuration Interaction (VMRCI) [12], has been able to identify relevant sub-blocks of the Hamiltonian matrix thereby reducing the size of the system. Similar construction was used with Adaptive-VCI (A-VCI) [13, 14] to access the smallest eigenvalues with a very good precision. These last approaches showed promising results in term of size reduction, but still require to a posteriori determine a big matrix block designed to improve the accuracy of the solutions. In the present work, the non zeros of this sub-block are never collected and the number of operation to perform proper Matrix Vector Products (MVPs) is highly reduced thanks to a new factorisation of the Hamiltonian. Beside, the expense of RAM is even more diminished because there is the possibility to constraint numerical accuracy on a few specified targets. In next sections are presented the context and the state of the art. The concept of duality intervenes all along the paper each time an association is made between a group of objects especially in section 4 explaining the theoretical aspects of the algorithm. Thereafter a description of the successive processed operations are followed by benchmarks. In this section, the method can provide comparable quality (i.e less than 1 cm^{-1} error) of a full A-VCI calculation, but with a memory consumption scaled down by more than a factor 15. Next, the list of input parameters of the program serves as a user manual. At the end, the conclusion also mentions a list of possible future developments.

2. Context

For a molecule composed of N_A atoms, one considers NM dimensionless normal coordinates [15] $\mathbf{q} = (q_1, q_2, q_3, \dots q_{\mathrm{NM}})$, with NM $= 3 * N_A - 6$. The corresponding harmonic frequencies are designated by $(\nu_1, \nu_2, \nu_3, \dots \nu_{\mathrm{NM}})$. The model is based on the Watson Hamiltonian [16] with zero rotational angular momentum (J=0). Its vibrational part contains the PES that is a multidimensional function known only for few points calculated by a first electronic resolution. It is represented in here as a Sum Of Product (SOP). For combinations of monomial degrees $\mathbf{c} \in \mathbb{N}^{\mathrm{NM}}$, attributed to force constants $K_{\mathbf{c}}$, the PES of maximal degree DP has the general form

$$\mathcal{U}_{\mathcal{K}}(\mathbf{q}) = \sum_{||\mathbf{c}||_1 \leqslant \mathrm{DP}} K_{\mathbf{c}} \prod_{n=1}^{\mathrm{NM}} q_n^{c_n}, \tag{1}$$

and the vibrational part writes

$$\mathcal{H}_{vib}(\mathbf{q}) = -\frac{1}{2} \sum_{n=1}^{\mathrm{NM}} \nu_n \frac{\partial^2}{\partial q_n^2} + \mathcal{U}_{\mathcal{K}}(\mathbf{q}). \tag{2}$$

The rotational part consists of the *Coriolis corrections*

$$\begin{aligned}
\mathcal{H}_{rot}(\mathbf{q}) &= \mathcal{C}_{ijkl}(\mathbf{q}) - \frac{1}{8} \sum_{\alpha=1}^{3} \mu_{\alpha\alpha}, \\
\mathcal{C}_{ijkl}(\mathbf{q}) &= -\frac{1}{2} \sum_{i<j} \sum_{k<l} Z_{ijkl} \left(\sqrt{\frac{\nu_j}{\nu_i}} q_i \frac{\partial}{\partial q_j} - \sqrt{\frac{\nu_i}{\nu_j}} q_j \frac{\partial}{\partial q_i} \right) \left(\sqrt{\frac{\nu_l}{\nu_k}} q_k \frac{\partial}{\partial q_l} - \sqrt{\frac{\nu_k}{\nu_l}} q_l \frac{\partial}{\partial q_k} \right) \\
Z_{ijkl} &= \sum_{(\alpha,\beta) \in (x,y,z)} \mu_{\alpha,\beta} \zeta_{ij}^{\alpha} \zeta_{kl}^{\beta},
\end{aligned} \tag{3}$$

where $\mu_{\alpha,\beta}$ is the inverse of the moment of inertia and coefficients $(\zeta_{ij}^{\alpha}, \zeta_{kl}^{\beta})$ are calculated according to the method of Meal and Polo [17]. The *Watson term* $-\frac{1}{8} \sum_{\alpha=1}^{3} \mu_{\alpha\alpha}$ modifies the eigenvalues by a constant so that the anhamornic frequencies calculated are not affected, then it will be omitted in the rest of the paper. In regard to the wave function of the total Hamiltonian $\mathcal{H}$, it is a linear combination of basis set elements belonging to an Ansatz B

$$\Psi(\mathbf{q}) = \sum_{\mathbf{b} \in B} x_{\mathbf{b}} \Phi_{\mathbf{b}}(\mathbf{q}), \tag{4}$$

each one writing as a product of one dimensional harmonic oscillators

$$\Phi_{\mathbf{b}}(\mathbf{q}) = \phi_{b_1}(q_1) \dots \phi_{b_n}(q_n) \dots \phi_{b_{\mathrm{NM}}}(q_{\mathrm{NM}}), \tag{5}$$

solution of the equation

$$\begin{aligned}
\mathcal{H}_0(\mathbf{q}) \Phi_{\mathbf{b}}(\mathbf{q}) &= E_{\mathbf{b}} \Phi_{\mathbf{b}}(\mathbf{q}), \\
E_{\mathbf{b}} &= \sum_{n=1}^{\mathrm{NM}} \nu_n (b_n + 1/2), \\
\mathcal{H}_0(\mathbf{q}) &= \sum_{n=1}^{\mathrm{NM}} \frac{\nu_n}{2} \left(-\frac{\partial^2}{\partial q_n^2} + q_n^2 \right).
\end{aligned} \tag{6}$$

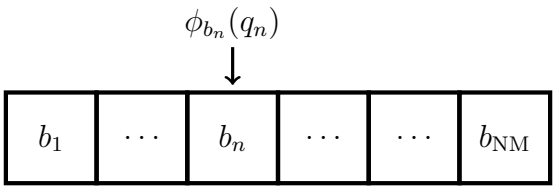

Figure 1: Multi index array identified with a basis function.
Each index is an Hermite function degree corresponding to an harmonic quantum level.

Implicitly, $\Phi_{\mathbf{b}}(\mathbf{q})$ is assimilated to the multi-index $\mathbf{b}$ (cf figure 1), and recognized by an integer coinciding with a pointer address when stored in memory. The algorithm could also possibly work with optimized basis set [4, 5], but the efficiency will be impoverished because DVCI fully uses harmonic oscillator properties. Afterwards, classical variational formulation leads to the eigenvalue problem

$$H\mathbf{X} = E\mathbf{X}, \tag{7}$$

with matrix coefficients of H built from the integrals[1]

$$\begin{aligned} \langle \Phi_{\mathbf{s}} | \mathcal{H} | \Phi_{\mathbf{b}} \rangle \quad &= \int_{\mathbb{R}^{\mathrm{NM}}} \Phi_{\mathbf{s}}(\mathbf{q}) \left(\mathcal{H}_{vib} + \mathcal{H}_{rot} \right)(\mathbf{q}) \Phi_{\mathbf{b}}(\mathbf{q}) \\ \forall (\mathbf{s}, \mathbf{b}) &\in B \times B. \end{aligned} \tag{8}$$

The curse of dimensionality. The number of basis functions exponentially increases with the number of nucleus when adopting a brute force variational method. As example, for a 12-d normal coordinate system, the dimension of the configuration space will be 11^{12} for a maximal quantum level equal to 10 in each direction. To solve the associated eigenvalue problem, one needs to manipulate items with the same size, and a multiplication by 8 gives about 251 terabytes of memory for a double precision vector.

3. State of the Art

To significantly reduce memory usage, matrix entries may not be stocked and computed with pruning conditions [18, 19, 20]. It consists in finding proper weights function α_n calculated on harmonic frequency criteria, and a maximal quantum level $\bar{d}$ to define the VCI space

$$\mathrm{VCI}_{\boldsymbol{\alpha}}(\bar{d}) = \{ \mathbf{b} \in \mathbb{N}^{\mathrm{NM}}, \sum_{n=1}^{\mathrm{NM}} \alpha_n(b_n) \leqslant \bar{d} \}. \tag{9}$$

For example $\alpha_n(b_n) = \lfloor \nu_n / \nu_{min} + 0.5 \rfloor * b_n$ [21]. When possible, symmetry properties may also be employed to separate basis set into groups of functions belonging to different irreducible representations. New and prospering works on tensorial factorisation relying upon alternating least square minimisation [22] permit a drastic reduction of RAM and time expenditure. In the present context, we can notice the efficacy of the Hierarchical Reduced-Rank Block

[1] Cf Appendix.

Power Method (HRRBPM) [23, 24] and the tensor train factorisation[25]. However, in this case the efficiency directly depends on the number of terms of the PES that should be small enough to observe accurate results with a low computational cost.
In the field of basis selection techniques, one widely uses combination of VCI and perturbation criteria with variation-perturbation theory [26, 27, 28, 29, 30, 31, 32, 33, 34]. For an member $\Phi_{\mathbf{b}}$ of a growing subspace B, selected basis functions $\Phi_{\mathbf{s}}$ should verify the perturbation criterion

$$\left| \frac{\langle \Phi_{\mathbf{s}} | \mathcal{H} - \mathcal{H}_0 | \Phi_{\mathbf{b}} \rangle^2}{E_{\mathbf{b}} - E_{\mathbf{s}}} \right| > \varepsilon_{\text{VP}}, \tag{10}$$

where ε_{VP} is a given threshold depending on the accuracy one wants to reach. $(E_{\mathbf{s}}, \Phi_{\mathbf{s}})$ and $(E_{\mathbf{b}}, \Phi_{\mathbf{b}})$ are eigen-pairs of $\mathcal{H}_0$ relying upon the nature of the basis functions that are used. It typically designs the harmonic part (6) or the sum of single-mode VSCF operators [35, 36, 37, 38].
An other relevant criteria for subspace selection is the residue. It is widely adopted in Davidson like methods [39, 40, 41, 42] and has recently been implemented in A-VCI [13, 14] sharing same structure than the A_k decomposition [3]. In the A_k theory, one considers primary and secondary spaces respectively called B and B_{S}. In the whole space $B \oplus B_{\text{S}}$, Hamiltonian matrix writes:

$$H = \left(\begin{array}{c|c} H_{\text{B}} & H_{\text{SB}}^T \\ \hline H_{\text{SB}} & H_{\text{S}} \end{array} \right), \tag{11}$$

where the different sub-blocks combine B and B_{S} in the following way

$$\begin{array}{rcl} [H_{\text{B}}] & = & [\langle \Phi_{\mathbf{s}} | \mathcal{H} | \Phi_{\mathbf{b}} \rangle]_{(\mathbf{s},\mathbf{b}) \in B \times B}\,, \\ [H_{\text{SB}}] & = & [\langle \Phi_{\mathbf{s}} | \mathcal{H} | \Phi_{\mathbf{b}} \rangle]_{(\mathbf{s},\mathbf{b}) \in B_{\text{S}} \times B}\,, \\ [H_{\text{S}}] & = & [\langle \Phi_{\mathbf{s}} | \mathcal{H} | \Phi_{\mathbf{b}} \rangle]_{(\mathbf{s},\mathbf{b}) \in B_{\text{S}} \times B_{\text{S}}}\,. \end{array} \tag{12}$$

For the A-VCI B_{S} is included in $\mathcal{H}_{vib}(B) \backslash B$, the complement of B in its image by $\mathcal{H}_{vib}(2)$, and for VMRCI $B_{\text{S}} \subset \text{STDQ}(B) \backslash B$ where STDQ refer to "Simple", "Double", "Triple" and "Quadruple" excitations. One can remark the use of the inclusion instead of equality because additional truncation on secondary space may be added to diminish the memory usage. For example one can cut down the maximal Harmonic energy as for the A-VCI, or only consider STD excitations in the case of VMRCI. The A-VCI is dedicated to compute the first eigenvalues of the Hamiltonian whereas VMRCI is focussed on any given initial state generally the fundamentals, first overtones, or combination bands. For an eigenpair $(E, \mathbf{X})$ of H_{B} (11), and $\tilde{\mathbf{X}} = (\mathbf{X}, \mathbf{0}_{B_{\text{S}}})$ the zero padded array of $\mathbf{X}$ in $B \oplus B_{\text{S}}$, the residual vector writes

$$\mathbf{Y}_{\text{S}} = H\tilde{\mathbf{X}} - E\tilde{\mathbf{X}} = (\mathbf{0}_B, H_{\text{SB}}\mathbf{X})^T. \tag{13}$$

It might easily be related to the correction vector $\mathbf{X}_\mathrm{s}$ obtained in the A_k theory[2]

$$\mathbf{X}_\mathrm{s} = (\mathbf{0}_B, (E.I_\mathrm{s} - \mathrm{diag}(H_\mathrm{s}))^{-1} H_{\mathrm{SB}}\mathbf{X})^T. \tag{14}$$

In an iterative process, A-VCI and VMRCI both consider maximal components of vectors $\mathbf{Y}_\mathrm{s}$ and $\mathbf{X}_\mathrm{s}$ to enrich primary space B. Since the non null coordinates of $(\mathbf{Y}_\mathrm{s}, \mathbf{X}_\mathrm{s})$ are always assigned to secondary space B_s, the chosen ones will be nullified at next iteration. The expected effect is to minimize the errors between eigenpairs computed in B and the ones that we would have calculated in $B \oplus B_\mathrm{s}$.

4. Theory and algorithm

In this part, we will see that the general algorithm is made from an enhanced mix of the previously recalled methods and a new theoretical approach to the construction of the Hamiltonian. Here are the main features:

- For the generation of B_s, in addition to the usual ways of expansion previously reminded, the method is capable to calibrate an optimal choice of excitations according to the analysis of the force field and rotational terms.
- The precision is focused on a given choice of targets, which allows an effective compression of the information and a significant gain of memory.
- The presented factorisation relieves the occupation of memory by performing on-the-fly operations for the MVP $H_{\mathrm{SB}}\mathbf{X}$ (13) that are not compensated by a systematic augmentation in execution time.

4.1. The local factorisation

Considering by $\hat{q}_n^+$ and $\hat{q}_n^-$ the creation (or raising) and annihilation (or lowering) operators[43, 44], acting on Hermite functions in the following manner

$$\begin{aligned} \hat{q}_n^+ \left|\phi_{b_n}\right\rangle &= \sqrt{b_n + 1}\left|\phi_{b_n+1}\right\rangle, \\ \hat{q}_n^- \left|\phi_{b_n}\right\rangle &= \sqrt{b_n}\left|\phi_{b_n-1}\right\rangle, \end{aligned} \tag{15}$$

the position and derivative may equivalently be written

$$\begin{aligned} q_n &= \frac{1}{\sqrt{2}}(\hat{q}_n^- + \hat{q}_n^+), \\ \frac{\partial}{\partial q_n} &= \frac{1}{\sqrt{2}}(\hat{q}_n^- - \hat{q}_n^+)\ . \end{aligned} \tag{16}$$

The local factorisation intends to answer the question:
Are there common factors in front of each product of excitations once any derivative or polynomial has been expressed in term of operators (16) and incorporated into $\mathcal{H}$?[3]

[2] $\mathrm{diag}(H_\mathrm{s})$: Diagonal of matrix H_s. $E.I_\mathrm{s}$: scalar E multiplied by identity matrix I_s.

[3] The product is non commutative, but to apprehend the problem, one needs to picture it as if it was.

We can, in fact, treat it from another equivalent angle by noticing from elementary properties of Hermite functions [4] that the paired elements $\langle\Phi_{\mathbf{s}}|\mathcal{H}_{vib}|\Phi_{\mathbf{b}}\rangle \neq 0$ involve only the local force constants[5]

$$\mathrm{LFK}(\mathbf{s}-\mathbf{b}) = \left\{ \begin{array}{l} \mathbf{c} \in \mathbb{N}^{\mathrm{NM}}, K_{\mathbf{c}} \neq 0, \\ \forall n \in \{1,\ldots,\mathrm{NM}\},\ K_{\mathbf{c}} \neq \nu_n/2, \\ \exists t_n \in \mathbb{N},\ |s_n - b_n| = c_n - 2t_n \end{array} \right\}. \tag{17}$$

In the same way, for a group of four canonical vectors $(\mathbf{1}_i, \mathbf{1}_j, \mathbf{1}_k, \mathbf{1}_l)$ of lengths NM with entry 1 respectively on position i, j, k, l, the local *Coriolis* interactions $\langle\Phi_{\mathbf{s}}|\mathcal{H}_{rot}|\Phi_{\mathbf{b}}\rangle \neq 0$ contain the combinations

$$\mathrm{LCI}(\mathbf{s}-\mathbf{b}) = \left\{ \begin{array}{l} \mathbf{c}^{ijkl} = \mathbf{1}_i + \mathbf{1}_j + \mathbf{1}_k + \mathbf{1}_l, \\ (i,j,k,l) \in \{1,\ldots,\mathrm{NM}\}^4, \\ i<j,\ k<l, \forall n \in \{1,\ldots,\mathrm{NM}\}, \\ \exists t_n \in \mathbb{N},\ |s_n - b_n| = c_n^{ijkl} - 2t_n \end{array} \right\}. \tag{18}$$

These sets are gathered in the local force field written

$$\mathrm{LFF}(\mathbf{s}-\mathbf{b}) = \mathrm{LFK}(\mathbf{s}-\mathbf{b}) \cup \mathrm{LCI}(\mathbf{s}-\mathbf{b}), \tag{19}$$

and defined no matter the sign of the differences $(s_n - b_n),\ n \in \{1,\ldots \mathrm{NM}\}$. With definitions (17),(18) combined to basic sorting and searching routines applied on all the monomial degrees $\{(\mathbf{c}, \mathbf{c}^{ijkl}),\ K_{\mathbf{c}} \neq 0,\ i<j,\ k<l\}$, the program builds the set of natural numbers associated to the non void local force fields[6]

$$\mathrm{LFF}^* = \left\{\mathbf{e} \in \mathbb{N}^{\mathrm{NM}},\ \mathrm{LFF}(\mathbf{e}) \neq \varnothing\right\} \cup \{\mathbf{0}_{\mathrm{NM}}\}. \tag{20}$$

From expressions (16), a creation is always accompanied by annihilation. Then, the dual of $\mathcal{H}$ consisting in its intrinsic sum of product of excitations, is constructed from the positive and negative multi-increments

$$\pm\mathrm{LFF}^* = \left\{(\pm e_1, \ldots, \pm e_{\mathrm{NM}}) \in \mathbb{Z}^{\mathrm{NM}},\ (e_1,\ldots,e_{\mathrm{NM}}) \in \mathrm{LFF}^*\right\}, \tag{21}$$

and writes[7]

$$\mathcal{H}^* = \sum_{\mathbf{e}\in\pm\mathrm{LFF}^*} \prod_{n=1}^{\mathrm{NM}} \hat{q}_n^{\pm e_n}. \tag{22}$$

The spaces B, B_{s} are growing up together along the repetitions of the main loop. One denotes by A the set of added basis functions in B from one iteration to the other. B_{s} is completed by browsing the image $\mathcal{H}^*(A)$. In the meantime the MVP $H_{\mathrm{SB}}\mathbf{X}$ is partially calculated for

[4] Cf Appendix.

[5] $K_{\mathbf{c}} \neq \nu_n/2$ because it is included in the harmonic part (6) and will be accounted when $\mathbf{s} = \mathbf{b}$.

[6] The zero excitation is systematically included.

[7] $\pm e_n$ is abusively employed to indicate the presence of a + or − sign in front of e_n.

the couples $(\mathbf{b}, \mathbf{b}+\mathbf{e}) \in (A \times \mathcal{H}^*(A) \backslash B)$ as [8]

$$
\begin{aligned}
(\widetilde{H_{\text{SB}}\mathbf{X}})_{\mathbf{b}+\mathbf{e}} \;=\; & \sum_{\mathbf{b}\in A} \Bigg[\sum_{\mathbf{c}\in\{\text{LFK}(\mathbf{e})\}} K_{\mathbf{c}} \prod_{n=1}^{\text{NM}} \langle \phi_{b_n+e_n} | q_n^{c_n} | \phi_{b_n} \rangle \\
& + \sum_{(i,j,k,l)\in\{\text{LCI}(\mathbf{e})\}} \langle \Phi_{\mathbf{b}+\mathbf{e}} | \mathcal{C}_{ijkl} | \Phi_{\mathbf{b}} \rangle \Bigg] x_{\mathbf{b}}.
\end{aligned}
\tag{23}
$$

The other part of the sum ($\mathbf{b} \in B \backslash A$) and other components are then completed as explained in the next section. When setting the parameter DoGraph$\diamond$23 to zero, one also has the possibility to directly compute the whole vector $H_{\text{SB}}\mathbf{X}$ by fetching $\mathcal{H}^*(B)$ instead of $\mathcal{H}^*(A)$ in (23). In that case it takes additional execution time because many tests are required to locate the addresses of the members of B_{s}. However, it is balanced by a smaller usage of RAM.
In supplement to the classical truncations, the excitations of $\mathcal{H}^*$ can be selected with ThrKX$\diamond$9, defining a threshold on the sum of the force constants contained in each local force field. The purpose is to avoid the runaway of B_{s} size and incorporate only the most contributive basis functions to the residue.

4.2. Complementary storage

Solving an eigenvalue problem usually requires a significant amount of MVPs, then the non null coefficients of H_{B} (12) might rather be collected than evaluated on the fly. So far, for any related method, the complexity to determine a non null matrix entry is at least of order $O\left(\text{NM} * \text{NPES}\right)$, where NPES designates the total number of force constant in the PES. When using the local force field for the evaluation of $\langle \Phi_{\mathbf{s}} | \mathcal{H} | \Phi_{\mathbf{b}} \rangle$, it is enough to fetch

$$
(\mathbf{s}-\mathbf{b})^+ = (|s_1 - b_1|, \ldots, |s_{\text{NM}} - b_{\text{NM}})|) \text{ into } \text{LFF}^*(20), \tag{24}
$$

calculate the terms in bracket equation (23) for $\mathbf{e} = \mathbf{s} - \mathbf{b}$ and add the harmonic energy $E^0_{\mathbf{b}}$ when $\mathbf{b} = \mathbf{s}$. The localization of the position of $(\mathbf{s}-\mathbf{b})^+$ in LFF^* is performed with a binary search [45] costing $O(\text{NM} * \log(|\text{LFF}^*|))$ operations. Consequently, the total complexity is

$$
O\left(\text{NM} * \left[|\text{LFF}(\mathbf{s}-\mathbf{b})| + \log(|\text{LFF}^*|)\right]\right), \tag{25}
$$

and one can easily check that $|\text{LFF}(\mathbf{s}-\mathbf{b})| + \log(|\text{LFF}^*|) << \text{NPES}$. Indeed for $\mathbf{e} \neq \mathbf{0}_{\text{NM}}$[9], the cardinal of $\text{LFF}(\mathbf{e})(19)$ is always very much lower than NPES. The worst case being NPES/2*NM happening only for a PES with no crossings. It therefore appears that the local factorisation consumes less operations than traditional methods for the determination of matrix coefficients, and for a small amount of MVPs, will be capable to integrate into a calculation on the fly without jeopardizing the execution time. Nevertheless, it will be all the more accelerated as the address of the paired elements is known in advance. This technique is actually employed for the residual block when the parameter DoGraph is strictly positif,

[8] $\mathcal{C}_{ijkl}$ is defined equation (3).
[9] The case $\mathbf{e} = \mathbf{0}_{\text{NM}}$ intervenes only for the construction of the diagonal elements of H_{B}.

by keeping only the pointers on the non zeros of H_{SB}, thus directly accessible to complete the MVP (23) for the missing parts:

$$(H_{\text{SB}}\mathbf{X})_{\mathbf{s}} = (\widetilde{H_{\text{SB}}\mathbf{X}})_{\mathbf{s}} + \sum_{\mathbf{b}\in(B\backslash A)} \langle\Phi_{\mathbf{s}}|\mathcal{H}|\Phi_{\mathbf{b}}\rangle\, x_{\mathbf{b}},\ \forall \mathbf{s} \in \mathcal{H}^*(B)\backslash B. \tag{26}$$

4.3. Initial space construction

Let's consider the ordered eigenvalues of block H_{B} built at iteration i

$$E_0^{\text{i}} \leqslant \ldots \leqslant E_\ell^{\text{i}} \leqslant E_{\ell+1}^{\text{i}} \leqslant \ldots. \tag{27}$$

One can demonstrate with the Poincaré separation theorem[46] that each E_ℓ^{i} is a decreasing sequence of i. Consequently, the minimization of the differences

$$E_\ell^{\text{i}} - E_\ell^{\text{i}+1} \tag{28}$$

could be effective only if no spectral hole is introduced in the interval holding targets at step zero where the eigenvalues are simple harmonic energies (6). Also, in order to integrate the perturbation effects, the initial space writes

$$B = \{\mathbf{b} \ /\ E_{\mathbf{b}} \leqslant E_{max} * \kappa\}, \tag{29}$$

where [10] $E_{max} = \text{MaxFreq} + E_{\mathbf{0}}$, MaxFreq⋄17 is the maximal tracked frequency and κ ⋄18 an empirical elongation factor accounting the global deviation between converged eigenvalues and associated harmonic energies. Its default value is set to 1.2, but it is automatically increased in agreement with the anharmonicity growing up as we get away from the smallest eigenvalue. To avoid timing issues caused by the great number of combination band[11] to be tested, the initial space is calculated with a greedy algorithm consisting in recursively applying one modal raising excitation starting from $\mathbf{0}_{\text{NM}}$.

5. The iterative process

The maximal size of the arrays used for $B, B_{\text{S}}, H_{\text{B}}$ and H_{SB} is determined with the allocated memory controlled by the parameter Memory ⋄19. At any iteration, the set {Targ} refers to eigenvectors of H_{B} having one component larger than ThrCoor ⋄14 and assigned to targets given by the parameter TargetState ⋄13. After a prior construction of the objects $\{\text{LFF}(\mathbf{e}),\ \mathbf{e} \in \text{LFF}^*\}$(19), the sequence of successive main steps decomposes as

<1> Build the initial subspace (29).

<2> Compute the eigenpairs $(E_\ell, \mathbf{X}_\ell)$ of H_{B} with the Implicitly Restarted Lanczos Method of ARPACK [47, 48]. The upper limit of calculated eigenvalues is chosen with MaxEv ⋄24.

[10] $E_{\mathbf{0}} = \dfrac{1}{2}\sum_{n=1}^{\text{NM}} \nu_n$.

[11] $\sum_{n}^{\text{NM}} b_n\nu_n$.

<3> Evaluate the residual vectors on the fly $H_{\text{SB}}\mathbf{X}_\ell,\ \ell \in \{\text{Targ}\}$ and secondary basis set B_{s} as explained in section 4. The expansion of B_{s} can additionally be reduced with an elimination of the less contributing excitations of $\mathcal{H}^*$ via the parameter ThrKX$\diamond$9, by restricting the maximal harmonic energy with Freq0Max$\diamond$8 or the maximal quantum level MaxQLevel$\diamond$7. Biggest components (in absolute value) of residual vectors are then employed to select basis functions to be added for next iteration through the inputs NAdd $\diamond$10, EtaComp $\diamond$11 and MaxAdd $\diamond$12.

<4> Go back to step <2> as long as

$$\max_{\ell\in\{\text{Targ}\}} \frac{\|H_{\text{SB}}\mathbf{X}_\ell\|}{E_\ell} > \text{EpsRez} \quad \diamond 5.$$

This criterion can be fulfilled only if enough memory has been allocated at the beginning. Otherwise, the algorithm will stop until maximal number of basis functions has been reached. It also trivially appears that increasing the targets does also augment the required memory to make them converge at once.

6. Benchmarks

All the calculus are done on a 64 bits, 2.70GHz quad core processor (model Intel i5-3340M) with 8 Gigabytes of RAM. No parallel process is used in here. DVCI program runs on a single CPU so that the actual computational time is the same as the CPU wall time.

6.1. CH_3CN*: Acetonitrile*

Methods are compared on molecule CH_3CN with the same PES as the one used in Ref. [19, 49, 24, 13] that was initially introduced by Bégué and al. [50], computed at CCSDT/cc-pVTZ level for harmonic frequencies and B3LYP/cc-pVTZ for higher order terms. This PES counts 311 terms (12 quadratic, 108 cubic and 191 quartic). The benchmark results are taken from Avila et Carrington [19] where symmetry has been employed to separate the full VCI space into two smaller subspaces. This PES has a small number of non null derivatives regarding the size of the molecule. The reference space defined as in [19] is the pruned basis set

$$\left\{ \begin{array}{l} \mathbf{b} \in \mathbb{N}^{12},\ \sum_{n=1}^{12} \alpha_n b_n \leqslant 27, \\ \alpha_1 = 3, \alpha_2 = 4, \alpha_3 = 3, \alpha_4 = 3, \alpha_5 = \alpha_6 = 3, \\ \alpha_7 = \alpha_8 = 4, \alpha_9 = \alpha_{10} = 3, \alpha_{11} = \alpha_{12} = 1. \end{array} \right\} \tag{30}$$

It contains 743 103 basis functions and the fundamental harmonic frequencies are

$$\begin{aligned} \nu_1 &= 3\,065,\ \nu_2 = 2\,297,\ \nu_3 = 1\,413,\ \nu_4 = 920, \\ \nu_5 &= \nu_6 = 3\,149,\ \nu_7 = \nu_8 = 1\,487, \\ \nu_9 &= \nu_{10} = 1\,061,\ \nu_{11} = \nu_{12} = 361\,(\text{cm}^{-1}). \end{aligned}$$

Results.

Table 1: Acetonitrile anharmonic fundamental frequencies

Assignment (Component)	Frequency (number)	Relative Residue	Absolute error Here-Ref	Experimental values
$\nu_0(0.97)$	9837.43(0)	0.0015	?	
$\nu_{11}(0.97)$	361.11(1)	0.0034	-0.1198	362 [51]
$\nu_{12}(0.97)$	361.17(2)	0.0039	-0.1779	
$\nu_4(0.95)$	900.76(6)	0.0032	-0.1001	920 [51]
$\nu_9(0.97)$	1034.25(7)	0.0033	-0.1202	1041 [51]
$\nu_{10}(0.97)$	1034.35(8)	0.0041	-0.2229	
$\nu_3(0.74)$, $\nu_9+\nu_{11}(0.44)$, $\nu_{10}+\nu_{12}(0.44)$	1389.17(15)	0.0038	-0.1980	1385 [51]
$\nu_3(0.62)$, $\nu_9+\nu_{11}(0.52)$, $\nu_{10}+\nu_{12}(0.52)$	1397.94(17)	0.0042	-0.2485	1402 [52]
$\nu_7(0.97)$	1483.33(20)	0.0031	-0.1034	1450 [52]
$\nu_8(0.97)$	1483.46(21)	0.0040	-0.2341	
$\nu_2(0.90)$	2250.94(70)	0.0037	-0.1923	2267 [51]
$\nu_1(0.60)$, $2\nu_7(0.51)$, $2\nu_8(0.51)$	2947.42(187)	0.0043	-0.3665	
$\nu_1(0.67)$, $2\nu_7(0.45)$, $2\nu_8(0.45)$	2981.01(193)	0.0037	-0.2316	2954 [51]
$\nu_5(0.93)$	3049.12(218)	0.0039	?	3009 [51]
$\nu_6(0.93)$	3049.16(219)	0.0040	?	

Performances summary.

Table 2: Performances summary on Acetonitrile fundamental targets. The CPU wall time is in second with the total number of iterations showed in parenthesis. EtaComp=3(cf$\diamond$11), NAdd=50 (cf $\diamond$10).

Final size of B	Final size of B_s	Final $nnz(H_B)$	Final $nnz(H_{SB})$	CPU Wall time(s) (Iterations)	Memory usage (Mo)
6169	590840	247662	3213078	352(4)	132

The complexity of this problem comes from the fact that the fundamental anharmonic frequencies are dispatched far away from the extremity of the spectrum. In the recent work of Ondunlami et al[14], the A-VCI didn't catch up all of them when computing the first 238 eigenvalues. Their best declared results parallelized on a 24-core Intel Xeon E5-2680 processors running at 2.8 GHz shows a maximal absolute error (relatively to the same reference) equal to 0.305 cm^{-1} on the first 121 eigenvalues with a time equal 5637 seconds

and a final basis set of 86 238 elements. In the HRRBPM of Thomas et Carrington[24] all the anharmonic frequencies before 2209 cm^{-1} were computed with a error lower than 0.38 cm^{-1}, 3.2 hours cpu time and 139 MB on a single Intel Core i7-4770 processor running at 3.4 GHz. In here all the fundamentals are computed with 132 megabytes memory for a maximal error lower than 0.37 cm^{-1} and a time of 5 minutes 52 seconds on a single CPU running at 2.70GHz.

6.2. C_2H_4 *: Ethylene*

The potential energy surface is originally the sixth order curvilinear symmetry-adapted coordinates of Delahaye & al [53] transformed into the sextic normal coordinate force field with PyPES [54]. In there work, point symmetry group D2h of C_2H_4 , has been exploited to divide VCI matrix into 8 symmetry blocks of respective dimension 106 889, 101 265, 100 366, 105 518, 105 643, 101 145, 101 255, 105 697. No symmetry assumption is applied in here. Reference calculation is taken from PyPES benchmarks[54] and consists on the 8−degrees excitation space ($\sum_{n=1}^{12} b_n \leqslant 8$) containing 125 270 basis functions.

Harmonic Frequencies and derivative orders of the PES.

$\nu_1 : 825.0,\quad \nu_2 : 950.2,\quad \nu_3 : 966.4,$
$\nu_4 : 1050.8,\quad \nu_5 : 1246.8,\quad \nu_6 : 1369.4,$
$\nu_7 : 1478.5,\quad \nu_8 : 1672.6,\quad \nu_9 : 3140.9,$
$\nu_{10} : 3156.8,\quad \nu_{11} : 3222.9,\quad \nu_{12} : 3248.7.$

Derivative order	2	3	4	5	6
Number of terms	45	147	290	642	1732

The second derivatives contain the harmonic and non null crossed terms.

Results.

Table 3: Ethylene anharmonic fundamental frequencies

Assignment (Component)	Frequency (number)	Relative Residue	Absolute error Here-Ref	Experimental values [55]
$\nu_0(0.98)$	11017.05(0)	0.0035	?	
$\nu_1(0.98)$	823.71(1)	0.0053	-0.0116	825.93
$\nu_2(0.98)$	935.42(2)	0.0056	-0.1234	939.86
$\nu_3(0.98)$	950.85(3)	0.0057	-0.1523	948.9
$\nu_4(0.98)$	1026.03(4)	0.0055	-0.1310	1025.69
$\nu_5(0.98)$	1225.05(5)	0.0056	-0.1468	1222
$\nu_6(0.97)$	1342.91(6)	0.0045	-0.0083	1343.54
$\nu_7(0.98)$	1442.02(7)	0.0055	-0.2236	1442.47
$\nu_8(0.90)$	1625.48(8)	0.0053	-0.0774	1625.4
$\nu_9(0.86)$	2985.80(60)	0.0060	-0.2960	2983.937
$\nu_{10}(0.80)$	3019.05(63)	0.0056	0.1493	3022.03

$\nu_{11}(0.91)$	3079.66(65)	0.0056	-0.2618	3083.36
$\nu_{12}(0.93)$	3101.60(70)	0.0054	-0.2998	3104.89

Performance summary.

Table 4: Performances summary on ethylene fundamental targets. The CPU wall time is in second with the total number of iterations indicated in parenthesis. EtaComp=2(cf⋄11), NAdd=100 (cf ⋄10).

Final size of B	Final size of B_s	Final $nnz(H_B)$	Final $nnz(H_{SB})$	CPU Wall time(s) (Iterations)	Memory usage (Mo)
19498	3674267	7491787	66159422	2238 (6)	760

The CPU wall time is equal to 37 minutes and 18 seconds. Even if the number of normal coordinates is the same as in the previous system, the memory requirement and the CPU time are significantly increased due to the higher number of terms in the PES.

6.3. C_2H_4O *: Ethylene oxide*

Here is computed the fundamentals of a 15-d Hamiltonian system where the PES is the one of Bégué and al [33] calculated at the CCSD(T)/cc-pVTZ level for harmonic frequencies and B3LYP/6-31+G(d,p) for the other terms. Apart from the harmonic force constants, the PES contains 180 cubic and 445 quartic terms. The reference results are taken from the A-VCI[14] where final basis set contains 7 118 214 elements.

Harmonic frequency in cm^{-1}.

$$\nu_1 : 3117.9,\ \nu_2 : 1549.1,\ \nu_3 : 1300.1,\ \nu_4 : 1157.9,\ \nu_5 : 899.6,$$
$$\nu_6 : 3196.6, \nu_7 : 1176.0,\ \nu_8 : 1052.2,\ \nu_9 : 3109.5,\ \nu_{10} : 1512.3,$$
$$\nu_{11} : 1156.8,\ \nu_{12} : 850.2,\ \nu_{13} : 3211.3,\ \nu_{14} : 1175.0,\ \nu_{15} : 815.5,$$

Correspondence with the ones listed in [14]

$$\nu_1 \equiv \omega_{13},\ \nu_2 \equiv \omega_{11},\ \nu_3 \equiv \omega_9,\ \nu_4 \equiv \omega_6,\ \nu_5 \equiv \omega_3,$$
$$\nu_6 \equiv \omega_{14}, \nu_7 \equiv \omega_8,\ \nu_8 \equiv \omega_4,\ \nu_9 \equiv \omega_{12},\ \nu_{10} \equiv \omega_{10},$$
$$\nu_{11} \equiv \omega_5,\ \nu_{12} \equiv \omega_2,\ \nu_{13} \equiv \omega_{15},\ \nu_{14} \equiv \omega_7,\ \nu_{15} \equiv \omega_1.$$

Results.

Table 5: Ethylene oxide anharmonic fundamental frequencies for five groups of targets separated by an horizontal bar. For comparison purpose, the zero point energy has the same value than in [14] (i.e 12461.473 cm^{-1}).

Assignment (Component)	Frequency (number)	Relative Residue	Error Here-Ref
$\nu_0(0.98)$	12461.47(0)	0.0034	0.0010
$\nu_{15}(0.97)$	792.96(1)	0.0045	-0.3305
$\nu_{12}(0.96)$	822.19(2)	0.0041	-0.2802
$\nu_5(0.97)$	878.51(3)	0.0038	-0.2350
$\nu_8(0.97)$	1017.47(4)	0.0044	-0.3319
$\nu_4(0.96)$	1121.47(5)	0.0042	-0.2993
$\nu_{11}(0.97)$	1123.92(6)	0.0042	-0.2962
$\nu_{14}(0.97)$	1146.03(7)	0.0042	-0.3077
$\nu_7(0.97)$	1148.19(8)	0.0037	-0.2299
$\nu_3(0.94)$	1271.17(9)	0.0047	-0.3909
$\nu_{10}(0.97)$	1467.58(10)	0.0037	-0.2402
$\nu_2(0.94)$	1495.49(11)	0.0042	-0.3334
$\nu_9(0.64)$, $\nu_2+\nu_{10}(0.52)$	2906.77(95)	0.0041	-0.4562
$\nu_9(0.52)$, $\nu_2+\nu_{10}(0.63)$	2989.70(111)	0.0044	-0.4552
$\nu_1(0.45)$, $2\nu_{10}(0.62)$, $\nu_8+\nu_{11}+\nu_{15}(0.44)$	2916.94(99)	0.0040	-0.2640
$\nu_1(0.52)$, $2\nu_{10}(0.62)$	2952.86(103)	0.0047	-0.4654
$\nu_6(0.85)$	3025.71(116)	0.0036	-0.3801
$\nu_{13}(0.80)$	3037.31(118)	0.0038	-0.411

Performances summary.

Table 6: Performances summary on ethylene oxide. Each tracked state(s) are separated by an horizontal bar. The CPU wall time is in second with the total number of iterations showed in parenthesis. NAdd=100 for all targets except for ν_{13} and ν_6 where NAdd=300. EtaComp=3 in any case.

Target(s)	Final size of B	Final size of B_s	Final $nnz(H_B)$	Final $nnz(H_{SB})$	CPU Wall time(s) (Iterations)	Memory usage (Mo)
ν_{15},ν_{12} $\nu_5,\nu_8,$ $\nu_4,\nu_{11},$ $\nu_{14},\nu_7,$ ν_3,ν_{10}	83346	11182617	5396776	80959806	11823(7)	1471
ν_9	133128	11840777	13352241	126500214	21507(16)	1690.6
ν_1	121180	11788217	11420147	117896340	21430 (16)	1632.1
ν_6	84554	8374499	7341322	79908192	11299(13)	1130.5
ν_{13}	118895	11082511	10947374	112694505	21808(14)	1579.8
Total	541103	54268621	48457860	517959057	87867	7504

The total CPU wall time is then 1 day 24 minutes and 27 seconds, and the total memory usage is 7.504 Gigabytes. The maximal absolute error on eigenvalues doesn't go over 0.47 cm^{-1} giving a maximal relative error lower than $4 * 10^{-5}$. As a matter of comparison the reference calculation were done with a total memory usage of 128 gigabytes and 3 days time

on a 24 cores computer meaning that the CPU wall time is much larger. Less accurate results (4-5 cm^{-1} error on higher frequencies) are achieved with HRRBPM [23] with a memory usage of 14.6 gigabytes and a CPU wall time of 8.7 days.

6.4. C_3H_3NO : *Oxazole*

The PES was constructed using the the adaptive density-guided approach (ADGA) introduced by Sparta et al [56, 57, 58]. The force constants, equilibrium geometry and normal coordinates where extracted from Madsen et al [59]. In their work they describe the construction of oxazole PES at CCSD(T)/cc-pVTZ level for the one-mode part and MP2/cc-pVTZ for the two-mode part. The three-mode part is extrapolated from the two-mode surface using MP2/cc-pVTZ gradients. The number of terms is 146 for the one mode, 4786 for the the two modes and 4335 for the three modes couplings.

Harmonic frequencies.

$$\begin{aligned}
\nu_1 : 603.8\text{cm}^{-1},\ \nu_2 : 644.2\text{cm}^{-1},\ \nu_3 : 748.9\text{cm}^{-1},\\
\nu_4 : 832.2\text{cm}^{-1},\ \nu_5 : 849.7\text{cm}^{-1},\ \nu_6 : 901.8\text{cm}^{-1},\\
\nu_7 : 913.8\text{cm}^{-1},\ \nu_8 : 1071.8\text{cm}^{-1},\ \nu_9 : 1109.2\text{cm}^{-1},\\
\nu_{10} : 1106.9\text{cm}^{-1},\ \nu_{11} : 1181.2\text{cm}^{-1},\ \nu_{12} : 1263.5\text{cm}^{-1},\\
\nu_{13} : 1348.3\text{cm}^{-1},\ \nu_{14} : 1533.4\text{cm}^{-1},\ \nu_{15} : 1570.4\text{cm}^{-1},\\
\nu_{16} : 3275.6\text{cm}^{-1},\ \nu_{17} : 3286.0\text{cm}^{-1},\ \nu_{18} : 3309.5\text{cm}^{-1}.
\end{aligned}$$

Table 7: Oxazole anharmonic frequencies for fundamental targets separated by an horizontal bar.

Eigenvalue number	Frequency	Relative Residue	Assignment (component)	Experimental values [60]
0	12559.9032	0.0046	$\nu_0(0.95)$	12457.5
1	592.8494	0.0074	$\nu_1(0.95)$	607 (A'')
2	631.2242	0.0075	$\nu_2(0.95)$	647 (A'')
3	727.3345	0.0072	$\nu_3(0.93)$	750 (A'')
4	795.0331	0.0077	$\nu_4(0.90)$	830(A'')
5	827.7750	0.0070	$\nu_5(0.90)$	854 (A'')
6	884.4607	0.0074	$\nu_6(0.92)$	899 (A'')
7	894.7569	0.0071	$\nu_7(0.91)$	907 (A'')
8	1031.4634	0.0064	$\nu_8(0.90)$	1046 (A')
9	1063.2165	0.0067	$\nu_{10}(0.82)$	1078 (A')
10	1075.3896	0.0069	$\nu_9(0.82)$	1086 (A')
11	1123.9852	0.0066	$\nu_{11}(0.89)$	1139 (A')
13	1217.7686	0.0066	$\nu_{12}(0.94)$	1252 (A')
16	1302.4880	0.0078	$\nu_{13}(0.90)$	1324 (A')
24	1481.3806	0.0079	$\nu_{14}(0.86)$	1504 (A')
27	1521.2312	0.0074	$\nu_{15}(0.91)$	1537 (A')

585	3125.9955	0.0086	$\nu_{16}(0.76)$	3141 (A')
608	3146.9603	0.0098	$\nu_{17}(0.82)$	3144 (A')
618	3159.5778	0.0087	$\nu_{18}(0.81)$	3170 (A')

Performances summary.

Table 8: Performances summary on Oxazole molecule. Each screened states are separated by an horizontal bar. The CPU wall time is in second with the total number of iterations indicated in parenthesis. In both cases, NAdd=200, EtaComp=3 and Freq0Max=20000 (cf ⋄8)

Target(s)	Final size of B	Final size of B_{s}	Final $nnz(H_{\text{B}})$	Final $nnz(H_{\text{SB}})$	CPU Wall time(s) (Iterations)	Memory usage (Mo)
$\nu_0, \ldots, \nu_{15}$	145820	27468841	87344774	299972877	74665(10)	6020
$\nu_{16}, \nu_{17}, \nu_{18}$	143916	29724836	60788880	285617698	170658(19)	5600
Total	289736	57193677	148133654	585590575	245323	11620

The total cpu wall time is 2 days 20 hours 8 minutes 43 seconds. A significantly bigger computational time for the second group of targets $\nu_{16}, \nu_{17}, \nu_{18}$ principally comes from the additional number of iterations. To a lesser extent, there is also the constraint to calculate the eigenvalues starting from the extremities of the spectrum as in Lanczos algorithm. A Davidson eigensolver [40, 41] or polynomial filtering techniques [61, 62, 63, 42] could be more adapted.

7. Input parameters

Presentation. The key words are case insensitive and should start at the beginning of each line of the input file. No specific order of apparition is required. Comments are indicated with the symbols '/' or '@'. The potential energy file should also be present in the directory where is executed DVCI (cf PESType⋄3). A minimal input file looks like :

```
NMode 6                  /  Number of normal coordinates
PESType 1                /  PES type of coefficients
OutName N2H2             /  Extension name for output files.
PESName N2H2_PES.in      /  Name of the file of the PES.
Memory 80               /  Maximal allocated memory in mega octets.
```

Detailed list of the key words.

⋄1 NMode
Designs the number of mass weighted normal coordinates.
NMode $= 3 * N_A - 6$. Where N_A stands for the number of atoms of the molecule.

⋄2 DoRot
Indicate if rotational corrections should be added to the Hamiltonian.
In the PES file, the section to be filled should start and end with words
COORDINATES ENDCOOR and each different field must be separated by an exclamation mark.

- If = 0
  then $\mathcal{H} = \mathcal{H}_{vib}(\mathbf{q})(2)$. If no section COORDINATES is written in the PES file then DoRot is automatically set to 0.
- If > 0 and even
  then $\mathcal{H} = \mathcal{H}_{vib}(\mathbf{q}) + \mathcal{H}_{rot}(\mathbf{q})$ (3). The equilibrium geometry in Bohr, atomic masses in electron rest mass, and normal coordinate eigenvectors should be indicated in the section COORDINATES like in the following example where the values has been extracted from Madsen & al[59].

```
COORDINATES
! Equilibrium geometry (bohr)
+1.188227703663817e+00  +1.871902685046820e+00  +5.700041831365050e-16 /C
-1.366889823622856e+00  +1.691283181574289e+00  +5.673619333132633e-16 /C
-1.998417882238274e+00  -8.244795338089052e-01  -2.155218936400323e-16 /O
+2.684304163103964e-01  -2.022882150581233e+00  -6.432553562388595e-16 /C
+2.229151382700317e+00  -5.583543727400242e-01  -2.215072662048790e-16 /N
+2.352687276664275e+00  +3.539113911722189e+00  +8.792950593981537e-16 /H
+2.239397718366979e-01  -4.057044206724038e+00  -1.063440276037351e-15 /H
-2.901750851724292e+00  +3.020857567289711e+00  +7.990414399553623e-16 /H
! Masses(me)
21874.6618172 /C
21874.6618172 /C
29156.9456749 /O
21874.6618172 /C
25526.0423547 /N
1837.15264562 /H
1837.15264562 /H
1837.15264562 /H
! Mode0:    X                       Y                       Z
 +3.231683769903235e-17 +6.115580033151089e-17 +4.025196381179206e-01
 +2.743657157380967e-17 +4.590882496215692e-17 -5.186966592183442e-01
 +1.991732529600559e-17 +5.710133055001405e-17 +5.169293296045854e-01
 -7.257396408556236e-17 +7.324385585074392e-17 -1.938605777276555e-01
 -9.053243042419710e-17 +2.384482767132475e-18 -1.671562436764069e-01
 +3.416768858945461e-17 -1.546457080242301e-17 +2.188043144086224e-01
 -1.950243964435382e-17 -1.457271158617074e-16 -2.002715238252209e-01
 -5.645471816646615e-17 -3.099000577024135e-17 -3.849787518532105e-01
! Mode1
 -4.471096934999722e-17 -2.133222140486714e-16 +3.245666277578175e-01
 -7.195147510789707e-18 -1.776354539431898e-16 -3.130286914609319e-02
 -1.161675776799252e-16 -2.465033916085211e-16 -2.713644979495499e-01
 -7.304974442800625e-17 +1.130401139043093e-16 +4.555118862530116e-01
 +1.980788418374012e-17 -1.509309233138445e-16 -5.878836462126088e-01
 +3.871182981501824e-17 +3.894914167531566e-17 +2.788008333658181e-01
 -1.775227773837671e-18 +1.687750614789954e-16 +4.342956786268967e-01
 -9.998627049911543e-18 -1.727718015508009e-16 -2.443444074954252e-02
! Mode2
```

```
 -1.362124303938439e-16 +1.648858093263572e-17 +7.985191937307874e-02
 +9.223533160690561e-17 -1.520416449194739e-16 +4.427598081406853e-01
 -3.106447728140722e-16 -2.024361724665231e-16 -7.334352062590022e-02
 +3.248565849336364e-17 +2.181702164863730e-17 -1.148459584461199e-01
 +3.231452029113849e-17 +1.995473253650141e-16 -3.854511885546462e-02
 -1.704399688795554e-16 +6.014500621640033e-17 -3.252722203195085e-01
 +1.114007103292430e-16 +8.683057684777304e-18 +1.582562080820018e-01
 +3.567043855971248e-16 +2.236379927154288e-16 -8.041677585822095e-01
! Mode3
 -8.530348729807672e-17 +5.064459244391696e-16 +2.099386275547177e-01
 -2.860512641462178e-17 +2.849372440086422e-16 -1.949978814638483e-02
 +9.671777362477948e-16 +3.228924378351529e-16 -1.437578533863218e-01
 -3.535112014919970e-16 -6.871677835537049e-16 +4.816962590153900e-01
 -6.328554966609436e-16 +1.463685147682383e-16 -1.649351614432971e-01
 +3.642789510512844e-16 +2.417462616289285e-18 -4.324245782423197e-01
 -5.058760423068888e-17 -3.092952891160942e-16 -6.990373338704934e-01
 -6.384663222086937e-16 -5.923791764375996e-16 -3.242248209522854e-04
       .                      .                      .
       .                      .                      .
       .                      .                      .
ENDCOOR
```

In accordance with the Wilson method[15], the normal coordinates are built from a set of $3N_A - 6$ eigenvectors $(Q_{i_\alpha})_{\ i\in\{1,\ldots,N_A\},\ \alpha\in\{1,2,3\}}$ of the Hessian matrix

$$\left(\frac{1}{\sqrt{m_i}\sqrt{m_j}}\frac{\partial^2 \mathcal{U}_{\mathcal{K}}}{\partial x_{i_\alpha}\partial x_{j_\beta}}\right)_{(i,j)\in\{1,\ldots,N_A\}^2,\ (\alpha,\beta)\in\{1,2,3\}^2} \tag{31}$$

derived from mass weighted displacements (m_i mass of nucleus i)

$$\Delta y_{i\alpha} = \sqrt{m_i}\,(x_{i_\alpha} - Xeq_{i_\alpha})\,,\quad i \in \{1,\ldots,N_A\},\ \alpha \in \{1,2,3\},$$

at the equilibrium geometry $\mathbf{Xeq} \in \mathbb{R}^{3N_A}$. The corresponding eigenvalues are the harmonic frequencies.

- If > 0 and odd
  then $\mathcal{H} = \mathcal{H}_{vib}(\mathbf{q}) + \mathcal{H}_{rot}(\mathbf{q})$ and the non mass weighted normal coordinate eigenvectors
  $$(Q_{i_\alpha}/\sqrt{m_i})_{\ i\in\{1,\ldots,N_A\},\ \alpha\in\{1,2,3\}}$$
  should be written instead of the classical ones.

◇3 PESType
Format of data's for the SOP PES.

- If PESType = 0 then the force constants $K_{\mathbf{c}}$ are expressed for dimensionless normal coordinates ($q_n = Q_n/\sqrt{\nu_n}$) and supplied in cm^{-1}. Regarding the format of the PES it starts and ends with the key words
  FORCEFIELD ENDFF. For NM normal coordinates, NM integers should be shown before the actual value of the force constant:

```
FORCEFIELD
 2 0 0 0 0 0 , 664.213550943134237
 4 0 0 0 0 0 , 4.335282791860437
 6 0 0 0 0 0 , -0.471897107116644
 0 2 0 0 0 0 , 675.140549094012272
 0 4 0 0 0 0 , 7.072599090662686
 2 0 1 0 1 0 , -12.115726349955748
 2 0 1 0 2 0 , 1.215724777835937
 2 0 1 0 3 0 , 0.161839021650279
 2 0 1 0 1 2 , 0.800398391535361
 2 0 1 0 0 2 , 0.662431742378760
 2 0 0 1 0 0 , 11.492641524810505
 2 0 0 2 0 0 , 0.117652856544075
 2 0 0 3 0 0 , -0.819356186630299
 . . . . . ,       .
 . . . . . .       .
 . . . . . .       .
ENDFF
```

Here it means that the first term is the one in front of q_1^2 namely $\nu_1/2$,
and for the last showed line 2 0 0 3 0 0, we are dealing with the force constant $K_{2,0,0,3,0,0} = -0.819356186630299$ cm^{-1} in agreement with the monomial $q_1^2 q_4^3$.

- If PESType $\geqslant 1$ then the derivatives are provided in atomic units, and for a Taylor expansion around the equilibrium position we have the correspondence

$$K_{c_1 \ldots c_{\mathrm{NM}}} = \frac{1}{c_1! c_2! \ldots c_{\mathrm{NM}}! \prod_{n=1}^{\mathrm{NM}} \sqrt{\nu_n}^{c_n}} \frac{\partial^{c_1,\ldots,c_{\mathrm{NM}}} \mathcal{U}_{\mathcal{K}}}{\partial Q_1^{c_1} \ldots \partial Q_{\mathrm{NM}}^{c_{\mathrm{NM}}}} * \mathrm{HartreeToCM}, \qquad (32)$$

where HartreeToCM is a converting factor from Hartree to cm^{-1} and $\nu_n = \sqrt{\frac{\partial^2 \mathcal{U}_{\mathcal{K}}}{\partial Q_n^2}}$.
The format of PES file is the same as the one supplied by the PyPES[64] library namely:

```
FORCEFIELD
[0,0,0,0,3,4 , 1.86861408859e-11],
[0,0,0,0,4 , -3.47804520495e-09],
[0,0,0,0,4,4 , 3.2867316136e-10],
[0,0,0,0,5,5 , 3.53998391655e-10],
[0,0,1,1 , 6.40569999931e-09],
[0,0,1,1,1,1 , -3.21892597605e-11],
[0,0,1,1,1,5 , -4.73206802979e-12],
[0,0,1,1,2 , 1.16300104731e-11],
[0,0,1,1,2,2 , -5.81559507099e-12],
[0,0,1,1,2,3 , 9.06462603669e-12],
[0,0,1,1,2,4 , -2.30405700341e-12],
[0,0,1,1,3 , -1.89925822496e-10],
```

```
[0,0,1,1,3,3 , -1.6879336773e-11],
[0,0,1,1,3,4 , 1.08730049094e-11],
 . . . . . ,       .
 . . . . . .       .
 . . . . . .       .
ENDFF
```

The repetitions are to be associated with a derivative order when coordinates are numbered starting from zero. For example the first line means

$$\frac{\partial^4}{\partial Q_1^4}\frac{\partial}{\partial Q_4}\frac{\partial}{\partial Q_5}\mathcal{U}_{\mathcal{K}} = 1.86861408859 \ * \ 10^{-11} \ \text{ a.u.}$$

⋄4 ThrPES
Threshold for PES force constants or derivatives. Default value is 10^{-20}.

⋄5 EpsRez
For eigenvectors of H_{B} (12) $\mathbf{X}_\ell$, $\ell \in \{\text{Targ}\}$, it is the maximal accepted relative residue

$$\max_{\ell\in\{\text{Targ}\}} \frac{\|H_{\text{SB}}\mathbf{X}_\ell\|}{E_\ell}$$

before the algorithm stop. They are built from the MVPs (26). The default value is $6 * 10^{-3}$.

⋄6 ThrMat
Minimal allowed absolute value of coefficients of H_{B}. Default is 10^{-20}. The matrix coefficients are computed with the full operator $\mathcal{H} = \mathcal{H}_{vib} + \mathcal{H}_{rot}$. If DoRot=0, only $\mathcal{H}_{vib}$ will be considered.

⋄7 MaxQLevel
This is the common maximal quantum level for the whole space $B \oplus B_{\text{s}}$. It increases together with distances between nucleus in motion and then should carefully be chosen conforming to the spacial region where the potential energy is still correctly represented and has no more than one local minimum. Each level on normal coordinate n will be adjusted with Freq0Max⋄8 as followed:

$$\text{MaxQLevel}(n) = \lceil \frac{\text{Freq0Max}}{\nu_n} \rceil + 1$$

⋄8 Freq0Max
Maximal allowed harmonic energy beyond the zero harmonic energy [12]$E_{\mathbf{0}}$ in $B \oplus B_{\text{s}}$. Default value is 30000.

[12] $E_{\mathbf{0}} = \frac{1}{2}\sum_{n=1}^{\text{NM}} \nu_n$

◇9 ThrKX
In operator $\mathcal{H}^*(22)$ only the increments $\mathbf{e} \in \mathrm{LFF}^*$ verifying [13]

$$\sum_{\mathbf{c}\in\{\mathrm{LFK}(\mathbf{e})\}} |K_{\mathbf{c}}| + \sum_{(i,j,k,l)\in\{\mathrm{LCI}(\mathbf{e})\}} |Z_{ijkl}| > \mathrm{ThrKX},$$

will be acceptable to generate the secondary space B_{s} and residual vectors. It should be strictly positive. Default value is 1.

◇10 NAdd
It is the minimal number of basis functions per non converged target states to be added for next iteration. They are chosen from maximal components (in absolute value) of the residual vectors (26)

$$\{(H_{\mathrm{SB}}\mathbf{X}_\ell)_{\mathbf{s}},\ \ell \in \mathrm{NotConv},\ \mathbf{s} \in B_{\mathrm{s}}\}, \tag{33}$$

where NotConv designs the set of non converged eigenpairs

$$\mathrm{NotConv} = \{\ell \in \{\mathrm{Targ}\},\ \frac{||H_{\mathrm{SB}}\mathbf{X}_\ell||}{|E_\ell|} > \mathrm{EpsRez}\}. \tag{34}$$

To accelerate convergence, NAdd is multiplied by $\mathtt{i}+1$, where $\mathtt{i}$ designates the iteration number.

◇11 EtaComp
The new added basis functions are selected from all the components of the residual vectors (26) greater than

$$\frac{1}{\mathrm{EtaComp} * \mathrm{NNotConv}} \sum_{\ell\in\mathrm{NotConv}} \|H_{\mathrm{SB}}\mathbf{X}_\ell\|_\infty$$

where NotConv (34) and NNotConv respectively stand for the set of non converged tracked eigenpairs of H_{B} and its cardinal. EtaComp should be greater than one, this guaranty that at list one component per non converged residual vector will be picked up.

◇12 MaxAdd
Limit for number of basis functions to add at each iteration. Default value is 1000.

◇13 TargetState
It indicates the maximal component of the eigenvectors of H_{B} that should be assigned to the targets matching with a multi-index array (cf figure 1). Except for $\mathbf{0}_{\mathrm{NM}}$ symbolized by 0(1), only its non zeros should be indicated with the characters $d(n)$ separated by a comma, where d stands for the degree of the Hermite function and n the normal coordinate. Alternatively TargetState can be followed by the label 'Fund' if the targets are the fundamentals and the ground state i.e $1(n),\ n = \{1, \ldots, \mathrm{NM}\}$ and 0(1). If 0(1) is not part of the targets then the zero point energy should be provided in cm^{-1} via the parameter GroundState ◇16.

[13] Z_{ijkl} is defined equation (3)

◇14 ThrCoor
Any eigenvector coordinate of H_{B} bigger (in absolute value) than this threshold and assigned to one of the targets, will be integrated into the iterative process and have its residual vector (26) calculated. In output, will be showed only the assignments of components larger than ThrCoor.

◇15 AddTarget
Sometimes, different eigenvectors point to the same maximal components. Then the actual number of targets is bigger than the one specified by the user. So it allocates additional arrays to correct this increasing. The default value is 2.

◇16 GroundState
Zero point energy required when it is not calculated (i.e not part of the targets). It can also be adopted as reference to printout the anharmonic frequencies.

◇17 MinFreq, MaxFreq
Frequencies in cm^{-1} specified to make converge all the eigenvalues within the interval

$$[\text{MinFreq} + \text{GroundState}, \text{MaxFreq} + \text{GroundState}]$$

when no target is indicated. If MinFreq is greater than zero, the value of GroundState should be supplied, else it will be computed. If MaxFreq is not given in input, then it will be set to the maximal harmonic frequency of tracked states for the initial subspace construction and to Freq0Max◇8 afterwards. Default values are [-100,4000].

◇18 Kappa
Elongation factor for the construction of the initial subspace estimating the global deviation between converged eigenvalues and associated harmonic energies. Its default value is 1.2 but it is automatically augmented to 1.3 when maximal target frequency is greater than 3000 cm^{-1}.

◇19 Memory
Total allocated memory in megabytes for the slots occupied by the eigensolver, the matrices (H_{B}, H_{SB}), the multi indexes (B, B_{S}), the PES, the local force fields and corresponding positive increments $\{\text{LFF}(\mathbf{e}),\ \mathbf{e} \in \text{LFF}^*\}$(19). This value will be used to set up the upper limit of basis functions SizeActMax that is appraised taking into account the array shrinkage factors KNREZ◇21, KNNZ◇20 and KNZREZ◇22. The total effective usage of memory is generally slightly bigger (about 5%) due to local array allocation.

◇20 KNNZ
Sparsity factor for H_{B}. The maximal number of non zero coefficients in H_{B} will be

$$\text{NNZActMax} = \text{KNNZ} * \text{SizeActMax} * \text{NXDualHTrunc}.$$

Where NXDualHTrunc is the number of excitations in $\mathcal{H}^*$ after truncation with ThrKX◇9. Default value is 0.03. Should be in]0,1].

◇21 KNREZ
Multiplicative factor of the maximal size of the residual space

$$\text{SizeRezMax} = \text{KNREZ} * \text{SizeActMax} * (\text{NXDualHTruncPos} - 1).$$

Where NXDualHTruncPos-1 is the number of raising excitations in operator $\mathcal{H}^*$ after truncation with ThrKX◇9 (the first excitation being zero). Default value is 0.2. Should be in]0,1].

◇22 KNZREZ
Shrinking factor for the maximal number of pointers on the non zeros of H_{SB}

$$\text{NNZRezMax} = \text{KNZREZ} * \text{SizeActMax} * \text{NXDualHTrunc}.$$

Where NXDualHTrunc is the number of excitations in operator $\mathcal{H}^*$ after truncation with ThrKX◇9. KNZREZ will be settled to zero when DoGraph=0.

◇23 DoGraph

- •If= 0 The MVPs $H_{\text{SB}}\mathbf{X}_\ell,\ \ell \in \{\text{Targ}\}$ are fully calculated by browsing $\mathcal{H}^*(B)$ instead of $\mathcal{H}^*(A)$ in (23).
- •If> 0 The row indexes and column pointers of the coupled elements of H_{SB} are stored in CSC[14] format to complete $\widetilde{H_{\text{SB}}\mathbf{X}}_\ell,\ \ell \in \{\text{Targ}\}$ in (23) for the missing entries (26). This option necessary increases memory requirement. The expense is about 40% greater in memory and 40% smaller in CPU time compared with DoGraph=0.

Default value is 1.

◇24 MaxEV
Maximum eigenvalues to be computed when counted from the smallest one. This number is adjusted to the size of the initial subspace minus one when it is actually larger than the latter. The eigensolver uses the Mode 1 and option WHICH='LM' of ARPACK subroutine DSAUPD. The greatest magnitude eigenvalues are computed on the shifted matrix

$$H_{\text{B}}^{'} = H_{\text{B}} - \text{Shift} * I_{\text{B}},\ \text{Shift} = \sum_{n=1}^{\text{NM}} \left[\frac{1}{2} + \text{MaxQLevel}(n) * \nu_n \right],$$

where I_{B} designates the identity matrix. Default value is 30.

◇25 DeltaNev
Reduce the number of wanted eigenvalues as

$$\text{MaxEV} = \text{Min}(\text{MaxEV}, \text{MaxScreen} + \text{DeltaNev}),$$

where MaxScreen is the higher position of the targets that tends to decrease with iterations. The purpose is to lighten the computational effort on the eigensolver. Default value is 1000.

[14]Compressed Sparse Column

◇26 MAXNCV
This is the maximal number of Lanczos basis vectors generated at each iteration in DSAUPD subroutine. Default value is 2*MaxEv◇24.

◇27 Tol
Stopping criterion for the relative accuracy of the Ritz values in DSAUPD subroutine. Default value is 10^{-8}.

◇28 RefName
Name of the input text file holding a floating point number at the beginning of each line to compare with the final results. The printed error is the difference between one of this value and the closest calculated frequency. Then it should manually be corrected when this correspondence is not true.

◇29 Verbose
When non equal to zero, it allows to print additional informations such as intermediate CPU times, position of targets in initial space, the center of mass, the moment of inertia and the characteristics of the dual operator.

◇30 OutName
Extension for output file names created when PrintOut$\neq$ 0 (cf◇31).

◇31 PrintOut

- If $\neq$ 0: All the final basis set and the components of screened states with coinciding eigenvalue number in parenthesis will be printed in the file OutName-FinalBasis.out. The eigenvectors in binary format are saved in the file OutName-Vectors.bin. These informations could be employed to compute infrared or Raman intensities in the final basis set.
- If = 0: No additional output file is created.

Default value is 0.

8. Conclusion

In this work has been presented a new algorithm to track specific states of molecular spectrum approaching the variational limit with a minimal usage of memory. Harmonic oscillator properties together with second quantization formulation were adopted to build a novel assemblage of structures available for dynamic subspace enrichment. The resulting code has shown challenging performances and could obviously be applied for bigger systems that the ones studied in here. For future improvements, one could use the final basis set and wave functions to compute infrared and Raman intensities. Also remains the possibility to adapt the method for different implementations of internal coordinates already available in a software like TROVE [65].

Acknowledgments

I would like to thank professor T.Carrington for providing me the potential energy surface of ethylene oxide.

Appendix : Hermite function analytical formulas

In one dimension, Hermite functions verify

$$\int_{\mathbb{R}} \psi_b(q) q^{d_1} \frac{\partial^{d_2}}{\partial q^{d_2}} \psi_{b+e}(q) dq \neq 0 \text{ if } |e| = d_1 + d_2 - 2t, \tag{35}$$

as well as for the switched product $\frac{\partial^{d_1}}{\partial q^{d_1}} q^{d_2}$.
It is easily demonstrable with recurrence relations [66, 67]

$$\begin{aligned} \psi_b'(q) = \sqrt{\frac{b}{2}} \psi_{b-1}(q) - \sqrt{\frac{b+1}{2}} \psi_{b+1}(q) \ , \\ q \ \psi_b(q) = \sqrt{\frac{b}{2}} \psi_{b-1}(q) + \sqrt{\frac{b+1}{2}} \psi_{b+1}(q), \end{aligned} \tag{36}$$

and can directly be related to the definition of operators (16). For the coefficients

$$\langle [q^d] \rangle_{b,s} = \langle \psi_b(q) | q^d | \psi_s(q) \rangle, \ (b,s) \in \{0, \ldots, \mathrm{Dim}\}^2, \ d \geqslant 1,$$

the following property applies

$$\langle [q]^d \rangle_{b,s} = \langle [q^d] \rangle_{b,s} \ , \ \forall (b,s) \in \{0, \ldots, \mathrm{Dim} - d + 1\}^2.$$

where

$$[q] = \begin{bmatrix} 0 & \sqrt{\frac{1}{2}} & 0 & \cdots & \cdots & 0 \\ \sqrt{\frac{1}{2}} & 0 & \sqrt{\frac{2}{2}} & \ddots & \ddots & \vdots \\ 0 & \sqrt{\frac{2}{2}} & 0 & \sqrt{\frac{3}{2}} & \ddots & \vdots \\ \vdots & \ddots & \sqrt{\frac{3}{2}} & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 & \sqrt{\frac{\mathrm{Dim}}{2}} \\ 0 & \cdots & \cdots & 0 & \sqrt{\frac{\mathrm{Dim}}{2}} & 0 \end{bmatrix}$$

is the well known Jacobi matrix constructed with Hermite function recurrence relations (36).

[1] K. Christoffel, J. Bowman, Investigations of self-consistent field, scf ci and virtual state-configuration interaction vibrational energies for a model three-mode system, Chem. Phys. Lett. 85 (2) (1982) 220–224. doi:10.1016/0009-2614(82)80335-7.

[2] T. Thompson, D. Truhlar, Scf ci calculations for vibrational eigenvalues and wavefunctions of systems exhibiting fermi resonance, Chem. Phys. Lett. 75 (1) (1980) 87–90. doi:10.1016/0009-2614(80)80470-2.

[3] Z. Gershgorn, I. Shavitt, An application of perturbation theory ideas in configuration interaction calculations, International Journal of Quantum Chemistry 2 (6) (1968) 751–759. doi:10.1002/qua.560020603.

[4] J. Bowman, K. Christoffel, F. Tobin, Application of scf-si theory to vibrational motion in polyatomic molecules, J. Phys. Chem. 83 (8) (1979) 905–920. doi:10.1021/j100471a005.

[5] J. M. Bowman+, The Self-Consistent-Field Approach to Polyatomic Vibrations, Acc. Chem. Res 19 (12) (1986) 202–208. doi:10.1021/ar00127a002.

[6] O. Christiansen, Moller-Plesset perturbation theory for vibrational wave functions, Journal of Chemical Physics 119 (12) (2003) 5773–5781. doi:10.1063/1.1601593.

[7] K. Yagi, H. Karasawa, S. Hirata, K. Hirao, First-principles quantum calculations on the infrared spectrum and vibrational dynamics of the guanine-cytosine base pair, ChemPhysChem 10 (9-10) (2009) 1442–1444. doi:10.1002/cphc.200900234.

[8] I. Respondek, D. M. Benoit, Fast degenerate correlation-corrected vibrational self-consistent field calculations of the vibrational spectrum of 4-mercaptopyridine, Journal of Chemical Physics 131 (5). doi:10.1063/1.3193708.

[9] M. Herman, D. S. Perry, Molecular spectroscopy and dynamics: a polyad-based perspective, Physical Chemistry Chemical Physics 15 (25) (2013) 9970. doi:10.1039/c3cp50463h.

[10] R. Roth, J. Langhammer, Pade-resummed high-order perturbation theory for nuclear structure calculations, Physics Letters B 683 (4-5) (2009) 6. arXiv:0910.3650, doi:10.1016/j.physletb.2009.12.046.

[11] O. Christiansen, Vibrational structure theory: new vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties, Phys. Chem. Chem. Phys. 9 (23) (2007) 2942–2953. doi:10.1039/B618764A.

[12] F. Pfeiffer, G. Rauhut, Multi-reference vibration correlation methods, Journal of Chemical Physics 140 (6), and references therein. doi:10.1063/1.4865098.

[13] R. Garnier, M. Odunlami, V. L. Bris, D. Bégué, I. Baraille, O. Coulaud, Adaptive vibrational configuration interaction (A-VCI): A posteriori error estimation to efficiently compute anharmonic IR spectra, The Journal of Chemical Physics 144 (20) (2016) 204123. doi:10.1063/1.4952414.

[14] M. Odunlami, V. L. Bris, D. Bégué, I. Baraille, O. Coulaud, A-VCI: A flexible method to efficiently compute vibrational spectra, The Journal of Chemical Physics 146 (21) (2017) 214108. doi:10.1063/1.4984266.

[15] E. B. Wilson, J. C. Decius, P. C. Cross, Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra, McGraw Hill, New York London, 1955.

[16] J. K. Watson, Simplification of the molecular vibration-rotation hamiltonian, Molecular Physics 15 (5) (1968) 479–490. doi:10.1080/00268976800101381.

[17] J. H. Meal, S. R. Polo, VibrationRotation Interaction in Polyatomic Molecules. II. The Determination of Coriolis Coupling Coefficients, The Journal of Chemical Physics 24 (6) (1956) 1126. doi:10.1063/1.1742729.

[18] G. Avila, T. Carrington, Solving the Schroedinger equation using Smolyak interpolants, Journal of Chemical Physics 139 (13). doi:10.1063/1.4821348.

[19] G. Avila, T. Carrington Jr., Using nonproduct quadrature grids to solve the vibrational schrdinger equation in 12d, J. Chem. Phys. 134 (5) (2011) 054126. doi:10.1063/1.3549817.

[20] G. Avila, T. Carrington, Using a pruned basis, a non-product quadrature grid, and the exact Watson normal-coordinate kinetic energy operator to solve the vibrational Schrdinger equation for C2H4, Journal of Chemical Physics 135 (6). doi:10.1063/1.3617249.

[21] J. Brown, T. Carrington, Using an expanding nondirect product harmonic basis with an iterative eigensolver to compute vibrational energy levels with as many as seven atoms, Journal of Chemical Physics 145 (14). doi:10.1063/1.4963916.

[22] G. Beylkin, M. J. Mohlenkamp, Algorithms for numerical analysis in high dimensions 26 (6) (2005) 2133–2159. doi:10.1137/040604959.

[23] P. S. Thomas, T. C. Jr, An intertwined method for making low-rank , sum-of-product basis functions that makes it possible to compute vibrational spectra of molecules with more than 10 atoms 204110 (2016) 1–40. doi:10.1063/1.4983695.

[24] P. S. Thomas, T. Carrington Jr., Using nested contractions and a hierarchical tensor format to compute vibrational spectra of molecules with seven atoms, J. Phys. Chem. A 119 (52) (2015) 13074–13091. doi:10.1021/acs.jpca.5b10015.

[25] M. Rakhuba, I. Oseledets, Calculating vibrational spectra of molecules using tensor train decomposition 124101. doi:10.1063/1.4962420.

[26] G. Rauhut, Configuration selection as a route towards efficient vibrational configuration interaction calculations, J. Chem. Phys. 127 (18) (2007) 184109. doi:10.1063/1.2790016.

[27] Y. Scribano, D. Benoit, Iterative active-space selection for vibrational configuration interaction calculations using a reduced-coupling vscf basis, Chem. Phys. Lett. 458 (4-6) (2008) 384–387. doi:10.1016/j.cplett.2008.05.001.

[28] M. Neff, G. Rauhut, Toward large scale vibrational configuration interaction calculations, J. Chem. Phys. 131 (12). doi:10.1063/1.3243862.

[29] G. Rauhut, T. Hrenar, A combined variational and perturbational study on the vibrational spectrum of P2F4, Chemical Physics 346 (1-3) (2008) 160–166. doi:10.1016/j.chemphys.2008.01.039.

[30] I. Baraille, C. Larrieu, A. Dargelos, M. Chaillet, Calculation of non-fundamental ir frequencies and intensities at the anharmonic level. i. the overtone, combination and difference bands of diazomethane, h2cn2, Chem. Phys. 273 (2-3) (2001) 91–101. doi:10.1016/S0301-0104(01)00489-X.

[31] P. Carbonnière, A. Dargelos, C. Pouchan, The vci-p code: An iterative variation-perturbation scheme for efficient computations of anharmonic vibrational levels and ir intensities of polyatomic molecules, Theor. Chem. Acc. 125 (3-6) (2010) 543–554. doi:10.1007/s00214-009-0689-7.

[32] C. Pouchan, K. Zaki, Ab initio configuration interaction determination of the overtone vibrations of methyleneimine in the region 28003200 cm[sup 1], The Journal of Chemical Physics 107 (2) (1997) 342. doi:10.1063/1.474395.

[33] D. Bégué, N. Gohaud, C. Pouchan, P. Cassam-Chenai, J. Liévin, A comparison of two methods for selecting vibrational configuration interaction spaces on a heptatomic system: Ethylene oxide, J. Chem. Phys. 127 (16) (2007) 164115. doi:10.1063/1.2795711.

[34] M. Sibaev, D. L. Crittenden, Balancing accuracy and efficiency in selecting vibrational configuration interaction basis states using vibrational perturbation theory, Journal of Chemical Physics 145 (6). doi:10.1063/1.4960600.

[35] G. Chaban, J. Jung, R. Gerber, Anharmonic vibrational spectroscopy of glycine:testing of ab initio and empirical potentials, J. Phys. Chem. A 104 (44) (2000) 10035–10044. doi:10.1021/jp002297t.

[36] G. M. Chaban, J. O. Jung, R. B. Gerber, Ab initio calculation of anharmonic vibrational states of polyatomic systems: Electronic structure combined with vibrational self-consistent field, The Journal of Chemical Physics 111 (5) (1999) 1823–1829. doi:10.1063/1.479452.

[37] J. O. Jung, R. B. Gerber, Vibrational wave functions and spectroscopy of (h2o)n, n=2,3,4,5: Vibrational selfconsistent field with correlation corrections, J. Chem. Phys. 105 (23) (1996) 10332–10348. doi:http://dx.doi.org/10.1063/1.472960.

[38] T. K. Roy, R. B. Gerber, Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications., Physical chemistry chemical physics : PCCP 15 (24) (2013) 9468–92. doi:10.1039/c3cp50739d.

[39] E. Davidson, The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices, J. Comput. Phys. 17 (1) (1975) 87–94. doi:10.1016/0021-9991(75)90065-0.

[40] F. Ribeiro, C. Iung, C. Leforestier, A Jacobi-Wilson description coupled to a block-Davidson algorithm: An efficient scheme to calculate highly excited vibrational levels, Journal of Chemical Physics 123 (5) (2005) 0–10. doi:10.1063/1.1997129.

[41] F. Ribeiro, C. Iung, C. Leforestier, Calculation of highly excited vibrational levels: A prediagonalized davidson scheme, Chem. Phys. Lett. 362 (3-4) (2002) 199–204. doi:10.1016/S0009-2614(02)00905-3.

[42] Y. Zhou, Y. Saad, A ChebyshevDavidson Algorithm for Large Symmetric Eigenproblems, SIAM Journal on Matrix Analysis and Applications 29 (3) (2007) 954–971. doi:10.1137/050630404.

[43] O. Christiansen, Vibrational coupled cluster theory., The Journal of chemical physics 120 (5) (2004) 2149–2159. doi:10.1063/1.1637579.

[44] C. Cohen-Tannoudji, B. Diu, F. Laloë, Quantum mechanics, Quantum Mechanics, Wiley, 1977.

[45] L. F. Williams, Jr., A modification to the half-interval search (binary search) method, in: Proceedings of the 14th Annual Southeast Regional Conference, ACM-SE 14, ACM, New York, NY, USA, 1976, pp. 95–101. doi:10.1145/503561.503582.

[46] E. Ziegel, Matrix Differential Calculus With Applications in Statistics and Econometrics, Technometrics 31 (4) (1989) 501–502. doi:10.1080/00401706.1989.10488622.

[47] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Natl. Bur. Stand. 45 (4) (1950) 255. doi:10.6028/jres.045.026.

[48] R. B. Lehoucq, D. C. Sorensen, C. Yang, ARPACK Users ' Guide : Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods ., Communication 6 (1998) 147. doi:10.1137/1.9780898719628.

[49] A. Leclerc, T. Carrington, Calculating vibrational spectra with sum of product basis functions without storing full-dimensional vectors or matrices, J. Chem. Phys. 140 (17) (2014) 174111. doi:10.1063/1.4871981.

[50] D. Begue, P. Carbonniere, C. Pouchan, Calculations of vibrational energy levels by using a hybrid ab initio and dft quartic force field: Application to acetonitrile, J. Phys. Chem. A 109 (20) (2005) 4611–4616. doi:10.1021/jp0406114.

[51] T. Shimanouchi, Tables of molecular vibrational frequencies. Consolidated volume II, Journal of Physical and Chemical Reference Data 6 (3) (1977) 993–1102. doi:10.1063/1.555560.

[52] R. Paso, R. Anttila, M. Koivusaari, The Infrared Spectrum of Methyl Cyanide Between 1240 and 1650 cm1: The Coupled Band System $\nu 3$, $\nu 16$, and $\nu 7 + \nu 28$, Journal of Molecular Spectroscopy 165 (2) (1994) 470–480. doi:10.1006/jmsp.1994.1150.

[53] T. Delahaye, A. Nikitin, M. Rey, P. G. Szalay, V. G. Tyuterev, A new accurate ground-state potential energy surface of ethylene and predictions for rotational and vibrational energy levels, Journal of Chemical Physics 141 (10) (2014) 0–16. doi:10.1063/1.4894419.

[54] M. Sibaev, D. L. Crittenden, The PyPES library of high quality semi-global potential energy surfaces, Journal of Computational Chemistry 36 (29) (2015) 2200–2207. doi:10.1002/jcc.24192.

[55] R. Georges, M. Bach, M. Herman, The vibrational energy pattern in ethylene (12C2H4), Molecular Physics 97 (1-2) (1999) 279–292. doi:10.1080/00268979909482829.

[56] M. Sparta, M. B. Hansen, E. Matito, D. Toffoli, O. Christiansen, Using electronic energy derivative information in automated potential energy surface construction for vibrational calculations, Journal of Chemical Theory and Computation 6 (10) (2010) 3162–3175. doi:10.1021/ct100229f.

[57] M. Sparta, I.-M. Høyvik, D. Toffoli, O. Christiansen, Potential Energy Surfaces for Vibrational Structure Calculations from a Multiresolution Adaptive Density-Guided Approach: Implementation and Test Calculations, The Journal of Physical Chemistry A 113 (30) (2009) 8712–8723. doi:10.1021/jp9035315.

[58] M. Sparta, D. Toffoli, O. Christiansen, An adaptive density-guided approach for the generation of potential energy surfaces of polyatomic molecules, Theoretical Chemistry Accounts 123 (5) (2009) 413–429. doi:10.1007/s00214-009-0532-1.

[59] N. K. Madsen, I. H. Godtliebsen, O. Christiansen, Efficient algorithms for solving the non-linear vibrational coupled-cluster equations using full and decomposed tensors, The Journal of Chemical Physics 146 (13) (2017) 134110. doi:10.1063/1.4979498.

[60] C. Pouchan, S. Senez, J. Raymond, H. Sauvaitre, Étude expérimentale et théorique des vibrations moléculaires de l'isoxazole, Journal de Chimie Physique 71 (1974) 525–532. doi:10.1051/jcp/1974710525.

[61] D. Sorensen, C. Yang, Accelerating the Lanczos algorithm via polynomial spectral transformations, TR97-29, Dept. of Computational and Applied Mathematics, Rice University, Houston, TX 0047.

[62] H. O. Karlsson, Calculation of highly excited vibrational states using a Richardson-Leja-Davidson scheme, Journal of Chemical Physics 126 (8). doi:10.1063/1.2646409.

[63] H. Fang, Y. Saad, A filtered Lanczos procedure for extreme and interior eigenvalue problems, SIAM J. SCI. COMPUT. 34 (4) (2012) 2220–2246. doi:10.1137/110836535.

[64] M. Sibaev, D. Crittenden, PyPES Extensible Library Manual (2016).
URL https://sourceforge.net/projects/pypes-lib-ext/files/

[65] S. N. Yurchenko, W. Thiel, P. Jensen, Theoretical ROVibrational Energies (TROVE): A robust numerical approach to the calculation of rovibrational energies for polyatomic molecules, Journal of Molecular Spectroscopy 245 (2) (2007) 126–140. doi:10.1016/j.jms.2007.07.009.

[66] I. S. Gradshteyn, I. M. Ryzhik, Table of integrals, series, and products, seventh Edition, Elsevier/Academic Press, Amsterdam, 2007.

[67] G. Szegö, Orthogonal Polynomials, no. vol.˜23 in American Mathematical Society colloquium publications, American mathematical society, 1939.