JOINT TREATMENT OF AB INITIO AND EXPERIMENTAL DATA ON VIBRATIONAL SPECTROSCOPY WITH STABLE NUMERICAL METHODS: DATABASE APPROACH

Gulnara M. Kuramshina
Department of Physical Chemistry, Chemical Faculty,
Moscow State University
Moscow 119899, Russia
kuramshi@phys.chem.msu.ru

ABSTRACT

Regularized algorithms for finding molecular force field parameters developed as the joint treatment of quantum mechanical and experimental data within stable numerical methods are successfully applied to the receiving the so-called regularized quantum mechanical force constants.

INTRODUCTION

Molecular force fields provide very important information about molecular structure and molecular dynamics and may be determined within harmonic approximation from experimental data of vibrational (Infrared and Raman) spectroscopy as a result of solving so-called inverse vibrational problem. Rapid progress in quantum mechanical calculations of theoretical harmonic force fields provides new ways for more accurate interpretation of experimental data as well as new possibilities for development of empirical force field calculations. The latter are particularly important for the large size molecules for which accurate *ab initio* calculations are impossible, so that empirical methods based on a solving an inverse vibrational problem still remain the best available source of force field parameters.

Fast progress in the investigation of rather large nanomolecules needs in development of special approaches for solving inverse vibrational problems moving far beyond traditional unstable methods based on the mean square root procedures. Analysis of large molecular systems (when a force constant matrix F is constructed from previously evaluated force constants of model compounds) run across to difficulties of possible incompatibility of results determined by different authors and by means of different numerical methods within different approximations (force field models). These difficulties related to nonuniqueness and instability of the solution of inverse vibrational

problem as well as to incompatibility within harmonic model of experimental data available.

To overcome the difficulties the special numerical algorithms based on Tikhonov's regularization method have been proposed [1-3]. In practical terms molecular spectroscopists employ model assumptions arising from the classic theory of chemical structure, involving concepts of bonded and nonbonded interactions, bond orders, monotone changes of the physico-chemical properties in series of related molecules, and the preservation of properties of separate molecular fragments, all related to the general concept of transferability of force constants. We proposed [1-3] to formalize these model assumptions and use them in the force field calculations. Calculated within such approach molecular force constants are included in the developed database of molecular parameters.

MATHEMATICAL FORMULATION OF INVERSE VIBRATIONAL PROBLEM

The idea of the force field arises from the attempt to consider a molecule as a mechanical system of nuclei while all the interactions due to electrons are included in an effective potential function $U(q_1,...,q_n)$, where $q_1,...,q_n$ denote the $n=3N-6$ generalized coordinates of the N atomic nuclei of the molecule. The potential function minimum (with respect to nucleus coordinates) assigns the equilibrium geometry of the molecule and the second derivatives of the potential with respect to nucleus coordinates in the equilibrium

$$f_{ij} = \frac{\partial^2 U}{\partial q_i \, \partial q_j}\Big|_{eq} \quad (i, j = 1,...n)$$

constitute a positive definite matrix F determining all the molecular characteristics connected with small vibrations. The vibrational

frequencies (obtained from IR and Raman spectra) are the main type of experimental information on molecular vibrations. They are connected with the matrix of force constants by the eigenvalue equation

$$GFL = L\Lambda \qquad (1)$$

where Λ is a diagonal matrix consisting of the squares of the molecular normal vibration frequencies $\omega_1,...\omega_n$, $\Lambda = diag\left\{\omega_1^2,...,\omega_n^2\right\}$, and G is the kinetic energy matrix in the momentum representation. L is a matrix of normalized relative amplitudes.

If only the experimental frequencies of one molecular isotopomer is known, inverse vibrational problem of finding force constant matrix F reduces to the inverse eigenvalue problem; hence, when G is not singular it follows that as a solution of Eq. (1) one has any matrix F such as

$$F = G^{-1/2}\, C^*\Lambda\, C\, G^{-1/2} \qquad (2)$$

where C is an arbitrary orthogonal matrix (the asterisk denotes the transposed matrix).

While Eq. (1) is the main source of data determining the force constants, it is evident that (except diatomic molecules) the $n(n+1)/2$ parameters of F cannot be found uniqueness from the n frequencies $\omega_1,...\omega_n$. This has led, on the one hand, to attempts to use certain model assumptions concerning the structure of the matrix F, and, on the other hand, to introduce the additional experimental data. Within the approximation considered, the force field of a molecule does not depend on the masses of the nuclei, and hence for the spectra of m isotopic species we have, instead of Eq. (1), the system

$$\left(G_i F\right)L_i = L_i\Lambda_i, \cdots i = 1,2,...m. \qquad (3)$$

The additional information may be extracted also from ro-vibrational spectra (Coriolis constants), gas electron diffraction (mean square amplitudes), etc. where molecular constants are determined by force constant matrix F.

The mathematical relation between the molecule vibrational properties (Eqs. (1),(3), etc.) and its experimental display can be summarized in the form of a single operator equation

$$AF = \Lambda . \qquad (4)$$

Here $F \in Z\ \acute{I}\ R^{n(n+1)/2}$ (Z is a set of possible solutions) is the unknown force constant matrix (real and symmetrical), $\Lambda \in R^m$ represents the set of available experimental data (vibrational frequencies, etc.)

Let introduce the following norms in the Euclidian space:

$$\|F\| = \left(\sum_{ij}^{n} f_{ij}^2\ \right)^{1/2} \quad ; \quad \|\Lambda\| = \left(\sum_{k=1}^{l} \lambda_k^2\rho_k\right)^{1/2}$$

where $r_k > 0$ are the positive weights; f_{ij} are the elements of matrix F; l_k $(k= 1,...m)$ are the components of L.

It is well known that the operator Eq. (4) can have finite or infinite number of solutions or even no solutions at all. The latter case can result from the crudeness of the physical (and mathematical) model which does not take into account some essential points, such as vibrational anharmonicities. The possible errors in the right-hand side and in operator A, however small, may cause finite (and large) variations of the solution F.

These properties of Eq. (4) make clear that the mathematically formalized problem of force field calculation belongs to the class of so-called ill-posed problems [3]. For solving ill-posed inverse vibrational problem, we have proposed to use Tikhonov's regularization method [1-3]. An inverse vibrational problem is formulated as a problem of finding the so-called normal solution (or normal pseudo(quasi) solution in the case of incompatibility of input data) of a nonlinear operator equation (4).

The solution we are searching for is a matrix $F^a \in Z$ that reproduces experimental data within given error level and is the nearest in the Euclidian metric to some given matrix F^0. All necessary model assumptions (explicit and implicit) concerning the form of force field may be taken into account by a choice of some given *a priori* matrix of force constants F^0 and a preassigned set D of *a priori* constraints on values of the force constants. This set defines a form of matrix F in the framework of the chosen force field model (i.e., with specified zero elements, equality of some force constants, etc.). If no *a priori* data constrains the form of solution, then D coincides with the set Z.

Let Eq. (4) with the exact data (A and Λ) has the unique normal solution $\overline{F}$ (with regard to a certain a priori specified matrix F^0).

It should be noted that if problem (4) with the exact given data has a unique solution, it essentially coincides with $\overline{F}$. If Eq. (4) has no solutions, it is necessary to formulize it as the problem of finding a normal pseudosolution (or quasisolution) i.e. search for a solution of the following problem:

It is required to obtain
$$\overline{F}_n = arg\,min\left\|F - F^0\right\|,$$
$$F \in \left\{ F :\ F \in D,\ \left\|AF - \Lambda\right\| = \mu \right\}$$
where $\mu = \inf\left\|AF - \Lambda\right\|,\ F \in D$.

The regularizing algorithm in this case must provide approximation converging to the normal pseudosolution of Eq. (4).

Now suppose that instead of A and Λ we are given some approximate A_h and Λ_δ such that

$$\left\|A_h F - AF\right\| \leq \varphi(h, F),\quad \left\|\Lambda_\delta - \Lambda\right\| \leq \delta,$$

where h, δ are known errors, and the error $\delta>0$ of the right-hand side of Eq. (4) is determined by the experimental errors of measurement, and $j(h,\delta)$ is a continuous function satisfying the condition

$$\sup_{\|F\|<C}\ \varphi(h, F) \to 0 \ \text{ when } h \to 0$$
for any finite $C>0$.

Now we construct stable approximation F_{hd} to the normal solution $\overline{F}_n$ such that $F_{hd} \to\ \grave{}F_n$ when $(h,\delta) \to 0$. The algorithms satisfying these conditions are determined as Tikhonov regularizing algorithm or regularizing operator.

In the case when the problem (4) has more than one normal solution (or pseudosolution) the approximations F_{hd} may converge to the set of normal solutions $\overline{F}_n$ (in the sense of β-convergence).

In Ref. [1-3] there were proposed some regularizing algorithms demonstrated their high efficiency.

One of possible algorithms is based on the minimization of the Tikhonov functional

$$M^\alpha[F] = \left\|A_h F - \Lambda_\delta\right\|^2 + \alpha\left\|F - F^0\right\|^2 \qquad (5)$$

where the regularizing parameter $\alpha > 0$ is determined within the generalized discrepancy principle [4]. This method is valid also for the special estimate of operator A_h given by

$$\varphi(h, F) = h\left\|A_h F\right\|.$$

The specific problem arises when there is known only a set of fundamental frequencies for a single molecule. It was proved in [1], that in this case the inverse vibrational problem is stable in the Hausdorff metrics in relation to the perturbations of the experimental frequencies and equilibrium geometry. As a result the problem of the finding the normal (pseudo)solution is also stable (if there exists more than one normal solution it means a stability in the sense of β-convergence too). We can reformulate the problem mentioned above in the next way:

It is required to find the minimum of the functional

$$\left\|F - F^0\right\|^2$$

on the set of matrices satisfied exactly to the experimental frequencies. F^0 is a given matrix in the linear space Z.

For a case of single molecule the variety of force constant matrices is described as Eq. (2). A matrix C may be performed as the multiple of $n(n-1)/2$ simple rotations. We have proposed [5] to search a normal solution of the Problem 2 by means of the Monte-Carlo method.

CONSTRAINTS ON THE VALUES OF FORCE CONSTANTS BASED ON AB INITIO CALCULATIONS

In the Tikhonov regularizing procedure, one can increase the stability and accuracy of the calculated solution F^a by using

a) an extended set of experimental data (including microwave spectra or ED data, etc.);

b) an improved choice of the stabilizer matrix F^0;

c) an improved choice of the constraint set D.

As a particularly effective choice of stabilizer, it was proposed [5] to include *ab initio* quantum mechanical results, in particular F^0 matrix, in the regularizing procedure. This leads to the concept of regularized quantum

mechanical force field (**RQM FF**), defined as the force constant matrix that is nearest to the some given quantum mechanical matrix F^0 and reproduces experimental frequencies within given error level.

The correct choice of constraint set D is also extremely important. Physically stipulated limitations may either decrease a range of possible matrices F, or provide criteria for selecting a concrete solution from a set of tolerable ones. An incorrect choice of constraints may lead to increasing incompatibility of the inverse problem, eventually resulting in a pseudosolution having no physical meaning. A set of a priori constraints may arise from several types of limitations on force constant values, e.g. [7]:

1. a part of force constants may be stipulated on a priori grounds to be a zero;

2. a part of force constants may be stipulated to satisfy inequalities $a_{ij} \leq f_{ij} \leq b_{ij}$, where a_{ij}, b_{ij} - are certain known values;

3. some force constants may be stipulated to conform to be equal in a series of related molecules (or conformers);

4. the final solution may be stipulated to conform to the so-called scaled force constant matrix [8], which may also be considered as a kind of constraint.

As a rule, the assumed limitations on the values of force constants of polyatomic molecules cannot be strictly proved and sometimes different theoretical levels can result in different forms of force constant matrix for a given molecule. Nevertheless, numerical quantum mechanical results on molecular force fields perform practically unique information and provide useful guidance in choosing realistic force field models for different types of molecules.

DATABASE APPROACH

The organization of the database of force constants requires applying the uniform principle of choosing the generalized coordinates and corresponding force constants.

As a background within database approach we've chosen the redundant systems of internal coordinates. All the algorithms throughout the package allow internal coordinates to be redundant. Redundancy conditions are taken into account automatically. Note also that conversion from Cartesian to internal coordinates may be not unique if coordinates are redundant. In this case, software package allows two choices [6]:

a) to generate canonical matrix F (that is, with minimal rank sufficient to represent all vibrational degrees of freedom, $3N$-6 where N is number of atoms);

b) to generate matrix F with least off-diagonal norm (this is one of commonly used model assumptions).

Earlier we've studied the regularized quantum mechanical force fields of different organic and inorganic molecules and analyzed their relationship to force constants of model force fields. From such analyses, we may draw some general conclusions:

(1) It is clearly evident that many molecules certain theoretical force constants connected with so-called "remote" interactions are very close to zero [7]. Force constants related with "rotor-rotor" interactions are very small, especially in comparison with diagonal elements, and may be safely neglected as is commonly assumed in the modified valence force field model.

(2) Force constants of common functional groups in related compounds or conformational isomers are highly similar and in many cases confirm the purely empirical rule of transferability of force constants in related molecules and the possibility of describing distinct conformers with some force constants assumed equal. It was shown [7] that theoretical force constants of some functional groups (e.g. methyl group, nitro-group, etc.) exhibited such similarity in a series of organic molecules.

This fact also supports the possibility of formulating inverse vibrational problems for a series of related molecules (or conformers) in terms of in-pair equalities of certain force constants as it was proposed in Ref. [1]

For the large size molecules (nanomolecules, clusters etc.) the empirical force field calculations are still remaining the best available source on force field parameters. The next scheme in such kind calculations includes the next steps: a) preliminary quantum mechanical analysis of moderate size molecular systems chosen as key molecules; b) joint treatment of *ab initio* and experimental data on vibrational spectra with stable numerical methods [2]; c) organization of a database on force field parameters transferable in a series of related compounds; d) normal coordinate analysis of large molecular systems, prediction of the

fundamental frequencies and thermodynamic functions.

These calculations 2)-4) may be performed with our software package SPECTRUM. All possible assumptions concerning the force field models considered above and different statements of inverse problem were realized in our software package SPECTRUM [3] allowing calculations of molecules formed by up to 300-400 atoms. While solving the inverse vibrational problems, the regularization parameter is chosen in accordance with generalized discrepancy principle [4].

This package allows the use of different sets of generalized coordinates (redundant or independent) in either direct or inverse vibrational problems, using various selections of experimental data, imposing special restrictions on elements of the force constant matrix (or solving unrestricted problems), jointly calculating force constant matrices for two or more related molecules. These programs therefore provide a flexible framework for obtaining molecular force fields under a variety of theoretical and experimental conditions.

It should be noted that the package allows treating more than one molecule simultaneously (and additionally include the data on several isotopic species). This is a sensible approach when the model assumptions require equivalence of certain force matrix elements. When this kind of constraints specified, these elements are held equivalent throughout the process of optimization. This option is of special value for verifying transferability properties of force constants.

All constraints can be applied to matrices in any chosen systems of generalized coordinates (Cartesian, internal, symmetry coordinates, etc.), being they redundant or independent. Redundancy conditions are taken into account automatically. While solving the inverse vibrational problem the regularization parameter is chosen in accordance with generalized discrepancy principle [4].

Regularizing algorithms based on the joint treatment of *ab initio* and experimental data allow us to take into account all features of *ab initio* results and to obtain force constant matrices that are the "best" in a sense of chosen physical model. For this purpose, we have studied the structures, vibrational spectra and force fields for different classes of polyatomic molecules by *ab initio* **HF** and **MP2** methods as well as with different **DFT** (Density Functional Theory) approaches using standard Gaussian basis sets performed with the **GAUSSIAN 9X** packages, calculated the regularized force constants and included them in our database.

REFERENCES

1. I.V. Kochikov, G.M. Kuramshina, A.G. Yagola, Stable numerical methods of solving certain inverse problems of vibrational spectroscopy. *U.S.S.R. Comput. Maths. Math. Phys.*, **27**, p. 33 (1987).

2. I.V. Kochikov, G.M. Kuramshina, Yu.A. Pentin, A.G. Yagola. Regularizing algorithms for molecular force field calculations. *J. Molec. Struct., **272** , p.13 (1992).

3. I.V. Kochikov, G.M. Kuramshina, Yu.A. Pentin, A.G. Yagola. *Inverse problems of vibrational spectroscopy*. VSP, The Netherlands, Zeist, 1999, p. 297.

4. A.N. Tikhonov, A.S Leonov and A.G. Yagola, *Nonlinear Ill-Posed Problems*, Chapman and Hall, London, 1997, p. 385.

5. G.M. Kuramshina, A.V. Lobanov, A.G. Yagola. Searching normal solution of inverse vibrational problem by the method of Monte-Carlo. *J. Molec. Struct*., **348**, p.147 (1995).

6. G.M. Kuramshina, F. Weinhold, I.V. Kochikov, Yu.A. Pentin, A.G. Yagola, Joint treatment of ab initio and experimental data in molecular force field calculations with Tikhonov's method of regularization. *J. Chem. Phys.*, **100**, p. 1414 (1994).

7. G.M. Kuramshina, A.G. Yagola, A Priori Limitations in Polyatomic Molecule Force Field Calculations. *Russian Journal of Structural Chemistry*, 32(2), p. 221 (1997).

8. G. Fogarasi, and P. Pulay. Quantum Chemical Calculation of Force Constants and Vibrational Spectra. *J. Molec. Struct.* **141**, p. 145, (1996).