

APPLICATION OF THE GENETIC ALGORITHMS TO DATA MODELING IN THE CASE OF FLUORESCENCE SPECTROSCOPY

M. Buczkowski^{a,1}, *J. J. Fisz*^{b,2}

^a Department of Physics and Biophysics, University of Warmia and Mazury, Olsztyn, Poland

^b Institute of Physics, Nicolas Copernicus University, Toruń, Poland

The genetic algorithms (GAs) are applied to extract the values of parameters that describe the behaviour of the asymmetric rotor-like fluorescence probes embedded in the uniaxially ordered system. The capabilities of the GA are enhanced by an application of a simple first order derivative search for the values of the multiplicative factors at each stage of the evolutionary run. The convergence is improved with the help of a gradient-expansion search performed after the GA is done. An example concerning the synthetic data set demonstrates the potentiality of this approach.

Генетические алгоритмы (ГА) используются для извлечения значений параметров, описывающих поведение асимметричных вращательноподобных флуоресцентных образцов, введенных в одноосно-упорядоченную систему. Возможности ГА возрастают, если использовать простой поиск производной первого порядка для значений мультипликативных факторов на каждой стадии протекания эволюции. Сходимость можно улучшить с помощью метода градиентного поиска после использования ГА. Потенциальные возможности этого приближения показаны на примере, связанном с искусственным набором данных.

PACS: 02.60.-X; 87.64.Ni

PREFACE

The main goal of this work is to show the possibility of an utilization of the genetic algorithms to the data modeling when the fluorescence depolarization technique (see, for example, Refs. [1,2]) is used to investigate the structure of the ordered system. The special emphasis is laid on the modification of the GA, which leads to better adaptation of this optimization method to this particular purpose.

The fluorescence probes embedded in the ordered system may provide information about their surround because their rotational mobility and distribution are affected by the molecules of the hosting matrix. The angle- and time-resolved measurements of the polarized fluorescence can be arranged in order to yield the data set consisting of several fluorescence decays [3], number of which depends on the experimenter.

¹E-mail: marcin.buczkowski@uwm.edu.pl

²E-mail: jjfisz@phys.uni.torun.pl

In search of the quantitative description of the system, we are to minimize the $\chi^2(\mathbf{u})$ merit function, which is defined as follows:

$$\chi^2(\mathbf{u}) = \frac{1}{\nu} \sum_c \sum_{i=1}^N \frac{[I_m^c(i) - I_l^c(i, \mathbf{u})]^2}{I_m^c(i)}, \quad (1)$$

where ν is the number of degrees of freedom, the index i runs over the channels of the instrument and the index c runs over the independent decays. $I_m^c(i)$ is the measured number of counts in the i th channel and $I_l^c(i, \mathbf{u})$ is the same quantity, but calculated according to the assumed model and to the vector of the model parameters denoted as \mathbf{u} . The particular choice and number of these parameters depend on the physical model involved. The potential restricted angular diffusion model is very useful and applied along this work.

1. GENETIC ALGORITHMS — A VERY BRIEF OVERVIEW

The genetics algorithm is a method of optimization that mimics the process of the natural selection. Firstly introduced by J. H. Holland in Ref. [4], the GA treats the set of the model parameters as a genotype. The values of the parameters constitute a particular individual and the value of the optimized function describes its fitness. The population of the individuals evolves in order to find the fittest individual. There are mechanisms of the natural evolution, such as cross-over and mutation, which serve to obtain improved values of parameters, applied to the population of the individuals. A detailed description of the GA is beyond the scope of this work, and can be found in numerous publications, see, for example, Refs. [5,6]. The GA is mentioned as a global algorithm, i.e., at its start it needs only information about the ranges of the sought values of the parameters. It is a great advantage in comparison with a local (e.g., gradient-expansion) algorithm, whose convergence is very sensible to the choice of an initial guess. From the point of view of the numerical calculations, another advantage of the GA is that it does not need the values of the derivatives of the optimized function.

To obtain the result presented in this work, the PIKAIA [7] implementation of the GA was used. This implementation is known by a decimal way of encoding the parameters values.

2. THE GA SUPPORTED BY OTHER ALGORITHMS

Aside from the above-mentioned advantages, the GA has also several drawbacks. In some cases, the information that provide the derivatives of the minimized function is very easily accessible, and it can simplify considerably the search of the minimum. The plain GA does not utilize this possibility at all. The other flaw is that close to the final stage of the search, the GA is likely to move around the wanted minimum point in the parameter space. This can be eliminated by an application of some algorithm performing a local search, which makes use of the GA result as of the starting point. The latter modification gives rise to the so-called hybrid approach [7].

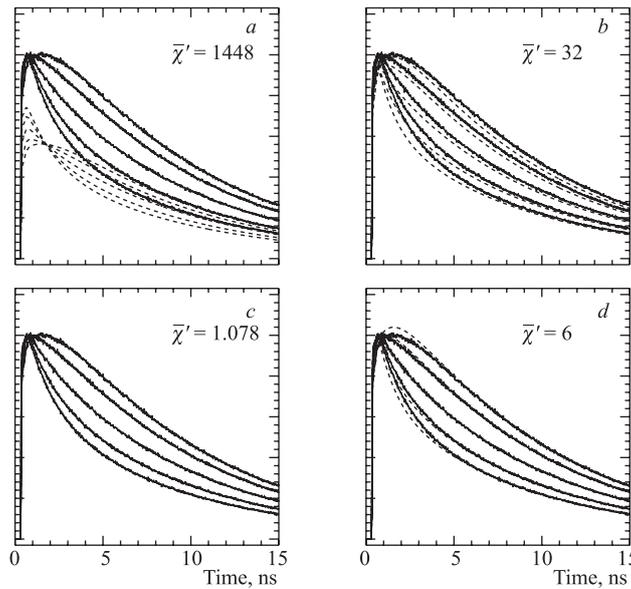
Let us expatiate on the first point. The formula for the fluorescence intensity decays can be partitioned in the following manner:

$$I_l^c(i, \mathbf{u}) = A_c \tilde{I}_l^c(i, \tilde{\mathbf{u}}), \quad (2)$$

where A_c are the positively valued constants and $\tilde{I}_i^c(i, \tilde{\mathbf{u}})$ are functions that are positively valued in the whole of their domain. Each function $\tilde{I}_i^c(i, \tilde{\mathbf{u}})$ depends on the set of parameters $\tilde{\mathbf{u}}$, which does not include A_c . The idea is to calculate «the best» A_c having some fixed values of $\tilde{\mathbf{u}}$, by means of equating the first partial derivative of the merit function (1) to zero, $\partial\chi^2(\tilde{\mathbf{u}}, A_c)/\partial A_c = 0$, which is very simple due to the fact that $\chi^2(\mathbf{u})$ is quadratic in A_c and leads to the following expression usable to calculate A_c [3]:

$$A_c = \frac{\sum_{i=1}^N \tilde{I}_i^c(i, \tilde{\mathbf{u}})}{\sum_{i=1}^N \frac{[\tilde{I}_i^c(i, \tilde{\mathbf{u}})]^2}{I_m(i)}}. \quad (3)$$

From the physical point of view, the amplitudes A_c are related to incidence conditions, a measurement geometry and a quantum yield of the fluorescence probes, but not to the investigated structure of the ordered system, quantified in the values of $\tilde{\mathbf{u}}$. In fact, this line of conduct is equivalent to scaling the fluorescence decays, or in other words, to state that the fluorescence decay with very different amplitude A_c , but obtained for the similar parameters $\tilde{\mathbf{u}}$, is better than that of similar amplitude but different $\tilde{\mathbf{u}}$, as depicted in the figure below.



The results of the scaling. The data set includes five fluorescence decays of different polarization. The dashed line curves are the calculated decays. *a)* The intensities that are distant from the modeled ones only due to different amplitudes; *b)* the intensities related to different values of $\tilde{\mathbf{u}}$ but with similar values of A_c ; *c), d)* the calculated intensities after applying the scaling. In panel *c* the fitted curves are covered by the modeled decays

Now, the improved values of A_c can be used to calculate the revised values of the fluorescence intensity that, in turn, become the subject of the further optimization.

The whole operation schematizes as follows:

1. Take \tilde{u}_{old} from the previous step of the GA and assume some values of A_c , say 1.

2. Calculate new A_c , according to Eq. (3).
3. Calculate the improved values of the fluorescence intensity.
4. Use the values of the calculated fluorescence intensity as the guidelines for the GA operations, in order to obtain new set of parameters \tilde{u} , \tilde{u}_{new} .
5. Set $\tilde{u}_{\text{old}} = \tilde{u}_{\text{new}}$ and go to 1.

3. APPLICATION OF THE GA TO ANALYZE THE POLARIZED FLUORESCENCE DECAYS

In order to model the results of the fluorescence depolarization measurements in the uniaxially oriented media we use the restricted rotational diffusion model. It describes Brownian rotational movements of the probes influenced by their surroundings, which give rise to the depolarization of the emitted light. In this formalism [8], the conditional probability distribution function $p(\Omega_0, 0 | \Omega, t)$ of finding the molecule at time $t > 0$ in the angular position Ω , if at time $t = 0$ its position is Ω_0 , is the solution to the equation of motion defined

by the time-independent time-development operator $\hat{H} = \sum_{i=1}^3 D_i \left\{ \hat{L}_i^2 + \hat{L}_i [\hat{L}_i V(\Omega)] / k_B T \right\}$,

where $\hat{L}_i, i = 1, 2, 3$ are the differential operators identical to the quantum mechanics angular momentum operators, k_B is the Boltzmann constant and T is the temperature. The way of solving this equation is described elsewhere [9]. The model parameters are the values of the coefficients u_{jn} of the aligning potential $V(\Omega)$ expansion in the base set of the Wigner functions $D_{m,n}^{(j)}(\Omega)$, $V(\Omega) = \sum_{j,n} u_{jn} D_{m,n}^{(j)*}(\Omega)$ and the elements of the diagonalized diffusion

tensor $D_i, i = 1, 2, 3$. The asymmetry of the rotator is assumed, so we must take into account that the three diagonal elements of the diffusion tensor are independent of each other and that the potential expansion coefficients u_{jn} with $n \neq 0$ appear.

The model parameters values and the best fit solution

Parameters	Simulation	Pikaia	Approximation	Errors
u_{20}	-1.6	-1.594	-1.594	± 0.094
u_{22}	-0.7	-0.690	-0.687	± 0.088
u_{40}	-0.9	-0.903	-0.90	± 0.15
u_{42}	-0.6	-0.594	-0.603	± 0.084
u_{44}	-0.4	-0.253	-0.24	± 0.44
$1/D_1$, ns	14	13.6	14.5	± 1.3
$1/D_2$, ns	12	12.0	11.87	± 0.13
$1/D_3$, ns	9	9.9	9.0	± 1.1

The whole set of the model parameters is given in the first column of the table. In the second column the parameters values used to obtain the synthetic data set are listed. The third column contains the best result of the nine runs of the GA. Each run comprises a population of a hundred of individuals evolving through 1200 generations. The result of each run served as a start point for the Marquadt method. The recovered values of parameters are listed in the fourth column. The best solution of the local searching algorithm origins from the

starting point provided by the best result of the GA. The local search procedure introduced some modifications to the parameters values, clearly visible in the case of the elements of the diffusion tensor. At last, the maximum errors estimations $\Delta \tilde{u}_i$ are collected in the fifth column.

4. DISCUSSION

By means of the presented algorithm the necessity of calculating the expected values of A_c is eliminated from the main loop of the GA. The advantage is twofold. First, the rate of convergence increases. Second, the number of measured fluorescence decays, which the number of A_c is up to, depends on the experimenter's needs and demands. It can vary due to the number of different chosen directions of the fluorescence polarization or to the number of the time windows. Hence, varies the dimensionality of the parameter space. The application of the presented algorithm makes the GA searched subspace of the parameter space dependent only on the number of the physical model parameters.

REFERENCES

1. *Werner T.* Modern Fluorescence Spectroscopy / Ed. E. L. Wehry. V. 2. N. Y.: Plenum Press, 1972.
2. *Lakowicz J.R.* Principles of Fluorescence Spectroscopy. Dordrecht; N. Y.: Kluwer Acad. Publ. /Plenum Press, 1999.
3. *Fisz J.J.* // Chem. Phys. Lett. 2005. V. 407, Iss. 1–3. P. 8–12.
4. *Holland J.H.* // J. Assoc. Comp. Math. 1962. V. 3. P. 297.
5. *Mitchell M.* An Introduction to Genetic Algorithms. 5th ed. The MIT Press, 1999.
6. *Haupt R.L., Haupt S.E.* Practical Genetic Algorithms. 2nd ed. Wiley Intersci. 2004.
7. *Charbonneau P., Knapp B.* A User's Guide to PIKAIA 1.0. NCAR Technical Note 418+IA. National Center for Atmospheric Research. Boulder, 1995.
8. *Risken H.* Fokker–Planck Equation. 2nd ed. Springer–Verlag, 1989.
9. *Fisz J.J., Buczkowski M.* Asymmetric Rotor-like Probes to Polarized Fluorescence Study of the Macroscopically Oriented Uniaxial Media. Numerical Solution. I (in preparation).