Modeling the optical properties of transparent and absorbing dielectrics by means of symbolic regression

Qingmeng Li, Demetrio Macias, Alexandre Vial

Research Article | Vol. 30, No. 23 / 7 Nov 2022 / Optics Express 41862
Optics EXPRESS

Modeling the optical properties of transparent and absorbing dielectrics by means of symbolic regression

QINGMENG LI, DEMETRIO MACIAS,* AND ALEXANDRE VIAL

Laboratory Light, Nanomaterials & Nanotechnologies – L2n, University of Technology of Troyes & CNRS EMR 7004, Troyes, France
*demetrio.macias@utt.fr

Abstract: In this contribution we explore the possibilities and limitations of symbolic regression as an alternative to the approaches currently used to characterize the dispersive behavior of a given material. To this end, we make use of genetic programming to retrieve, from either ellipsometric or spectral data, closed-form expressions that model the optical properties of the materials studied. In a first stage we consider transparent dielectrics for our numerical experiments. Next we increase the complexity of the problem and consider absorbing dielectrics, which not only require the use of complex functions to model their dielectric function, but also imply a supplementary constraint imposed by the verification of the causality principle.

1. Introduction

Although the idea of using optimization techniques [1] or neural networks [2] for the solution of inverse problems in optics is not new, the advent of high performance computing facilities and the recent advances in the field of Artificial Intelligence (AI) have led to a significant number of works concerning the application of these computational tools in Nano-Optics and Nanotechnologies [3–5]. In particular, the characterization of the optical properties of metals [6–8], absorbing dielectrics [9], biological structures [10–13] and more recently metamaterials [14] has specially attracted the attention of different research groups. Those references share a common thread, they make use of experimentally measured or numerically generated data to search for the parameters of an established well known dispersion model, whose algebraic form can be more or less complex depending on the materials to be characterized.

An alternative way to treat the particular inverse problem just described could be by means of Symbolic Regression (SR), a special regression scheme that can identify and predict the underlying mathematical relation between the elements of a data set. The main difference between AI-based approaches and SR is that the latter can be considered as a "glass-box" that provides a closed-form expression that represents the model of the problem studied. The former, on the other hand, can be seen as a "black box" that finds the relationship between the different elements of the data set but it does not provide any closed expression [14,15]. The potential of SR has been illustrated within the fields of dynamical systems [16,17], nonlinear optics [18,19] or in material sciences [20], to cite but few examples. Furthermore, in a recent work Udrescu and Tegmark [21] made use of a modified SR scheme to recover 100 equations from the Feynman Lectures on Physics [22] and they claimed an improvement of the state-of-the-art success rate from 15 to 90%. Notwithstanding the spectacular results reported in the references just cited, none of them treats complex functions, which are unavoidable when characterizing, for example, the dispersive behavior of absorbing dielectrics, metals or the materials that compose a multilayer biological structure structure. It seems thus natural to explore the possibilities and limitations of SR within this context.

#468110 https://doi.org/10.1364/OE.468110
Journal © 2022 Received 16 Jun 2022; revised 3 Sep 2022; accepted 6 Sep 2022; published 31 Oct 2022

Optics EXPRESS

The structure of this contribution is as follows: In Sect. 2, we outline the operating principles of SR. In Sect. 3, we first set the numerical parameters required for our simulations. Next, we present some typical results obtained when we search for the dispersion models of transparent and absorbing materials from either ellipsometric or spectral data. In the case of complex dispersion models we include the additional constraint of causality in our SR scheme. We give our concluding remarks in Sect. 4.

2. Symbolic regression

Although several numerical implementations of SR have been published over past decades, it is possible to distinguish two main approaches of completely different natures. One makes use of a population-based heuristic method known as Genetic Programming (GP) [23]. An example of the other approach, based on deterministic arguments, is known as Fast Function Extraction (FFX) [24].

Throughout this work we make use of a GP-based SR. Although the basic principles of GP can be found elsewhere [23,25], for the sake of completeness we will briefly describe them in the following paragraphs.

The first step in the SR process is, as illustrated in Fig. 1, the random generation of a set of functions f_{SR} that represent the physical model to be retrieved. The size μ of this initial population remains constant throughout the optimization loop. Also, the algebraic form of its elements is defined by the allowed operations that may be performed by each function f_{SR}.

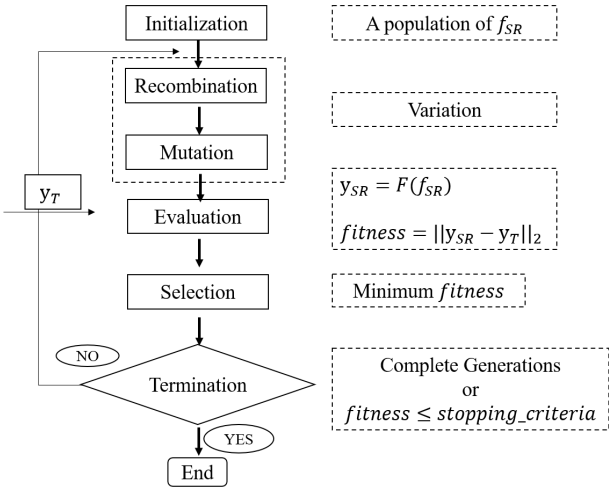


Fig. 1. Flux diagram with parameters of SR.

With reference to Fig. 1, the next step in the optimization loop is the variation of the elements in the initial population. This is done by means of the recombination and mutation operations, which depend on the so called hyper-parameters. These involve different probabilities of performing the genetic operations as, for example, the crossover probability P_{cr}, the subtree mutation probability P_{sm}, the hoist mutation probability P_{hm} and the point mutation probability P_{pm} [25]. Another important hyper-parameter, known as the parsimony pressure p_{co}, serves to control bloat, a well known and not yet completely solved problem of GP that is defined as the uncontrolled growth of the average size of an individual in the population [26]. It is noteworthy to say that the hyper-parameters must be set before the starting of the regression process and that there is not a fixed rule to establish them. However, as it will be mentioned in the next section, there are typical values well suited for the problems studied in this contribution.

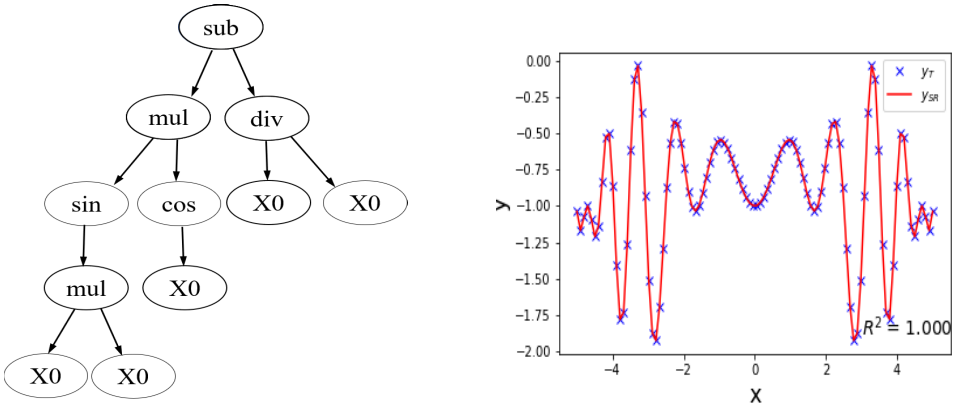
Optics EXPRESS

As shown in Fig. 1, the next stage in the optimization loop is the evaluation of each modified element. Although there exist several metrics, often the elements are evaluated trough the computation of the Euclidean norm, which measures the closeness between the target data y_T and the prediction y_{SR} obtained with the SR-based model. The Euclidean norm associates an objective or fitness value to each element. Only those elements in the population with a low fitness value will be selected for the next iteration of the optimization loop. Thus, when the fitness value is close to zero one may consider, at least in principle, that the closed-form expression that models the target data has been found.

To illustrate the operating principles of SR, we consider as reference the Nguyen's fifth benchmark function [27]

$$f_T(x) = \sin(x^2)\cos(x) - 1, \tag{1}$$

which in the current example serves also to generate the target data depicted with blue crosses in Fig. 2(b).



(a) (b)

Fig. 2. (a) A tree representation of the retrieved expression from GPLearn. (b) Comparison between the target and the predicted data computed through Nguyen F5 benchmark function (blue crosses) and Eq. (3) (solid red line), respectively. The related correlation coefficient is $R^2 = 1$.

Throughout this work we employ a GP-based SR numerical implementation known as GPLearn. This is an open source library coded in Python [28] that provides a closed-form expression in LISP format [29]

$$sub(mul(sin(mul(X0, X0)), cos(X0)), div(X0, X0)) \tag{2}$$

that can be written as

$$f_{SR}(X0) = \sin(X0^2)\cos(X0) - X0/X0, \tag{3}$$

and has a tree representation, as illustrated in Fig. 2(a). It is noteworthy to say that, in the context of GP, the length of an expression is defined as the number of functional and terminal nodes in the tree. Consequently, it is not necessarily related with the number of terms in the expression retrieved. In the current example, the length is 11 and not 2 as one might intuitively think. Also, it is evident that the term $X0/X0$ in the expression (3), known as intron, can be written as 1.

In Fig. 2(b) we compare the target data $y_T = f_T(x)$ with the data $y_{SR} = f_{SR}(x)$, depicted with a solid red line, computed by means of the expression retrieved.

Optics EXPRESS

Although it was not necessary for this simple example, in certain cases the closed-form expressions found require further simplification. This can be achieved by means of *Sympy*, another open source Python library devoted to symbolic calculus [30].

It is important to mention that to talk about local or global convergence within the frame of SR is not a trivial task. This is due to the fact that there is not a one-to-one relationship between the closed-form expressions found and their fitness value. It means that two expressions of different lengths and algebraic forms are equivalent. This can be an advantage for problems where there is not a well-established model, and one can choose the most suitable expression. However, in problems where the objective is to retrieve a known function, global convergence would be achieved only if all the expressions obtained from the search for the optimum, starting from different initial states, had the same length, the same fitness value and the same algebraic form as the searched function. However, this result is impossible to obtain because of the random nature of the recombination and mutation operators.

3. Results

At this stage some examples are convenient to assess the performance of our SR scheme when used to model the optical properties of a given material. Throughout this section, as illustrated in Fig. 3, we consider two case studies. In the first we search for the real dielectric function of a transparent dielectric. In the second case we extend the applicability of the SR scheme for the retrieval of the complex dielectric function of an absorbing dielectric. It is noteworthy to mention that in the presence of absorption the problem becomes even more difficult. This is because the SR scheme has to handle complex functions that must verify the causality principle.

As illustrated with a blue block in Fig. 3, regardless of the material, we consider two possible inputs to the regression scheme. The first are previously published ellipsometric data. To keep the complexity of the problem into a manageable level, the second possible input is spectral reflectance data generated when a flat interface, separating two semi-infinite media, is illuminated from vacuum with a plane wave at normal incidence. We assume that the transmission region is filled with a homogeneous, linear, isotropic and non-magnetic medium with complex refractive index $\tilde{n}(\omega) = n(\omega) + i\kappa(\omega)$. Then, the input reflectance spectrum is given by the equation [31]

$$R(\omega) = \left| \frac{1 - \tilde{n}(\omega)}{1 + \tilde{n}(\omega)} \right|^2 . \tag{4}$$

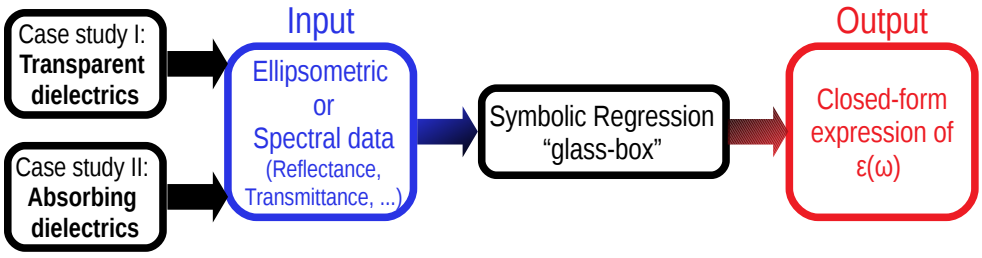


Fig. 3. Schematic of the application of SR to retrieve the dispersion model of a given material.

In any of the cases, the expected output, depicted with a red block in Fig. 3, must be a closed-form expression that models the dielectric function of a given material.

The results of extensive numerical experiments led us to set $g = 100$ as the number of iterations in the optimization loop. This parameter served also as termination criterion. Furthermore, the size of the population was set to $\mu = 5000$. The GP hyper-parameters were set to $P_{cr} = 0.7$ for

Optics EXPRESS

the crossover probability, $P_{sm} = 0.1$ for the sub-tree mutation probability, $P_{hm} = 0.1$ for the hoist mutation probability, $P_{pm} = 0.1$ for the point mutation probability and $p_{co} = 0.001 - 0.1$ for the parsimony coefficient. Moreover, we defined the functions set $f_s = \{'+','-','*','/',' ix'\}$, where $i = \sqrt{-1}$ and x is a real variable. The operation ix is used in the regression scheme only when searching for complex dielectric functions. Also, in each of our numerical experiments we search for the optimal solution from 50 different random initial states.

Throughout the numerical experiments conducted in this section the fitness function or lost error to be minimized is

$$f = \left\| \mathbf{F}^T - \mathbf{F}^{SR} \right\|_2, \tag{5}$$

where $\|...\|_2$ is the Euclidean norm, $\mathbf{F}^T$ may represent the target ellipsometric or spectral data, respectively ϵ or R, from where the dispersion model should be retrieved and $\mathbf{F}^{SR}$ are the data predicted by means of the closed-form expression found through the SR process.

In order to make an objective comparison between the target data and the prediction by the closed-expression retrieved, throughout this section we compute the mean absolute relative error (MARE), which has an intuitive interpretation in terms of the relative error [32]

$$MARE(y, y^T) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y_i^T}{y_i^T} \right|. \tag{6}$$

3.1. Retrieval of dispersion models of transparent dielectrics

In this first example, the target ellipsometric data were generated with Sellmeier's formula for SiO_2 [33,34]

$$\epsilon^T(\lambda) = 1 + \frac{0.6961663\lambda^2}{\lambda^2 - 0.0684043^2} + \frac{0.4079426\lambda^2}{\lambda^2 - 0.1162414^2} + \frac{0.8974794\lambda^2}{\lambda^2 - 9.896161^2}. \tag{7}$$

We show, in Fig. 4(a), the results obtained when we searched for the dielectric function from different initial states. Each point in the cloud represents the best closed-form expression obtained through the SR process. The x and the y axes respectively denote the length and the corresponding fitness value. The former is related to the algebraic form of the expression. The latter is a measure of the closeness between the target and the predicted data.

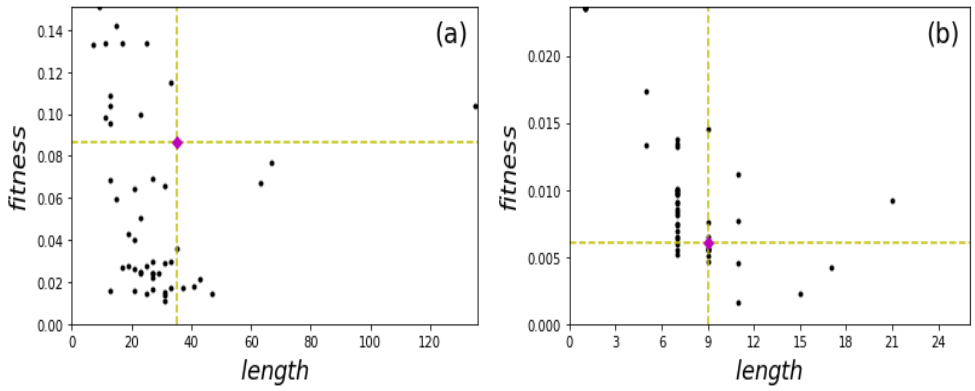


Fig. 4. Optimal solutions retrieved through the SR scheme from 50 different initial states. The pink diamonds represent the closed-form expression found considering the same initial state when the target data were generated: (a) with Eq. (7) and (b) with Eqs. (4) and (7).

As illustrated in Fig. 4(a), there is not a unique solution and it is then possible to find different closed-form expressions that model the material's dielectric function. However, throughout our

Optics EXPRESS

numerical experiments we found that best predictions are those of solutions with a moderate fitness and length values. A typical result of this situation is illustrated with the pink diamond in Fig. 4(a), whose related closed-form expression is given by

$$\epsilon_e^{SR}(\lambda) = 2.44838 + \frac{-2.277\lambda^4 + 0.6815\lambda^3 - 0.14648\lambda^2}{7.227\lambda^4 - 2.922\lambda^3 + 1.06193\lambda^2 - 0.1787\lambda + 0.038407}. \tag{8}$$

At a first sight, the algebraic form of Eq. (8) does not respect the parsimony principle and seems to be a direct consequence of bloat. In order to discuss this result, let us make use of Sellmeier's equation with two terms [35]

$$\epsilon(\lambda) = 1 + \sum_{i=1}^{2} \frac{B_i\lambda^2}{\lambda^2 - C_i}, \tag{9}$$

where B_i and C_i represent real coefficients. After some straightforward algebraic manipulations we arrive to the equation

$$\epsilon(\lambda) = 1 + \frac{\beta_1\lambda^4 - \beta_2\lambda^2}{\lambda^4 - \beta_3\lambda^2 + \beta_4}, \tag{10}$$

with $\beta_1 = B_1 + B_2$, $\beta_2 = B_1C_2 + B_2C_1$, $\beta_3 = C_1 + C_2$ and $\beta_4 = C_1C_2$.

Although the algebraic forms of Eqs. (9) and (10) are different and the physical interpretation of the second is less evident, both equations provide the same information. On this basis, a comparison of Eqs. (10) and (8) suggests that the algebraic form of Eq. (8) is not a failure of the SR scheme but just that it is unable to render the closed-form expression written in a compact form. Furthermore, this result can be interpreted as if the SR would have learned, during the optimization process, that it was necessary to add several corrective terms to the dispersion model to obtain an accurate fit. Then, contrary to common practice, increasing the parsimony pressure to search for a compact expression would have prevented the SR scheme from converging to the best solution in the current case.

In Fig. 5(a) we respectively depict with blue dots and red upside-down triangles the target dielectric function, computed with Eq. (7), and the dielectric function predicted by Eq. (8). Also, by means of Eq. (6) we find, $MARE\left[\epsilon_e^{SR}, \epsilon^T\right] = 0.001$ which confirms the acceptable agreement observed. Furthermore, in Fig. 5(b) we compare the target reflectance spectrum with the spectrum computed using Eq. (4). As expected , the agreement between the two curves is evident and corroborates that the retrieved closed-form expression found accurately models the dielectric function of SiO_2.

In order to explore further the capabilities of SR, we repeat the search for the dielectric function keeping the same conditions considered in the previous example as well as the same 50 initial states. However, this time $\mathbf{F} = R$ in Eq. (5) and the input to the SR scheme is a reflectance spectrum computed using Eq. (4) assuming $\Im\{\tilde{n}(\omega)\} = 0$.

The results obtained from spectral information are shown in Fig. 4(b). A somewhat expected result is the lack of uniqueness of the solution. Also, the scales of the axes differ significantly from those in Fig. 4(a). An interesting feature is that the solutions are not spread but rather confined to regions defined by specific lengths. The pink diamond in the figure corresponds to the same initial state shown in Fig. 4(a) and its related closed-form expression is

$$\epsilon_R^{SR}(\lambda) = 2.046666 + \frac{0.051}{\lambda}. \tag{11}$$

The significant difference between the algebraic forms of Eq. (11) and Eqs. (7) and (8) illustrates the effect of the input data on the SR-scheme. In this case the fitness value is related to the reflectance spectra and not, at least directly, to the dielectric function retrieved. Thus the problem is reduced to find the best fit between the spectra irrespective of whether the algebraic form of

Optics EXPRESS

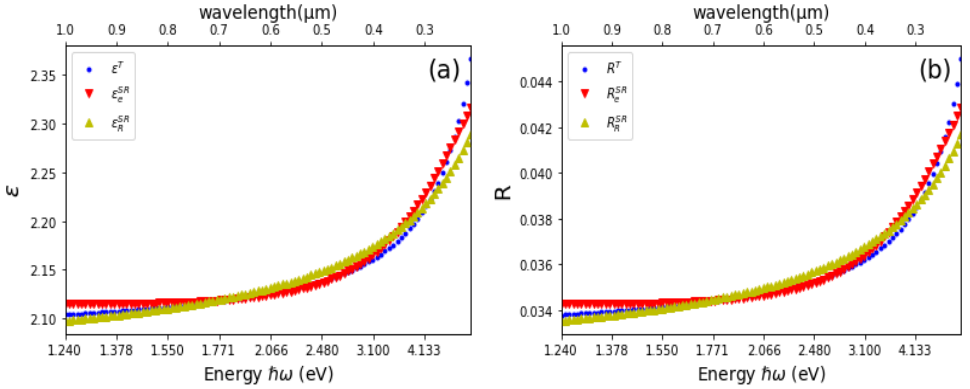


Fig. 5. (a) Target dielectric function generated with Eq. (7) (blue dotted line), the dielectric functions predicted by Eq. (8) (red upside down triangles) and Eq. (11) (yellow triangles). (b) Target spectrum (blue dots) and reflectance spectrum computed using Eqs. (11) and (4) (yellow triangles).

the expression found is that of the target dielectric function. This statement is supported by the comparison between the data predicted by Eq. (11), yellow triangles in Fig. 5(a), and the target data and confirmed by a $MARE\left[\epsilon_R^{SR}, \epsilon^T\right] = 0.004$. Furthermore, a visual comparison between the target spectrum and that generated considering Eq. (11) in Eq. (4) is shown in Fig. 5(b). The agreement between both spectra is also confirmed by a $MARE\left[R_R^{SR}, R^T\right] = 0.010$.

3.2. Retrieval of dispersion models of absorbing dielectrics

The results obtained in the previous paragraphs are encouraging and show the potential of SR for the case of lossless dielectrics. It seems thus natural to assess the performance of SR when absorbing dielectrics are considered. It is convenient to note that in this case the dielectric function is complex and this fact makes the search for the closed-form expression more difficult. This is because the real and the imaginary parts are not independent from each other but they must be related through Kramers-Kronig (K-K) relations in order to satisfy the principle of causality.

In this section, we proceed as we did for the case of transparent dielectrics and look for the dielectric function from 50 different initial states. Also, we keep the same hyper parameters of the SR-scheme. Moreover, once the closed-form expression has been found we compute K-K relations numerically using the implementation described in [36].

We consider Lorentz model for MgO with two oscillators as the target dielectric function [31]

$$\epsilon^T(\omega) = 3.01 + \frac{6.6 * 401^2}{(401^2 - \omega^2) - i7.62\omega} + \frac{0.045 * 640^2}{(640^2 - \omega^2) - i102.4\omega}, \tag{12}$$

where ω is in cm^{-1} and, in this example, $\omega \in [200, 800]$. That is, the frequencies considered are not in the visible range as it was the case for transparent dielectrics but in the far-infrared.

The results obtained when the dielectric function is searched from different initial states are shown in Fig. 6(a). The high fitness values together with the long lengths of several solutions give a clear indication of the complexity of the problem. As in the case of transparent dielectrics, the solution is not unique. However, those solutions with a low fitness value seem to be the best suited to model the complex dielectric function. The closed-form expression related to the pink diamond in Fig. 6(a) is given by

$$\epsilon_e^{SR}(\omega) = 4.046632 + \frac{\omega^2 - 2\omega + 1}{A}, \tag{13}$$

Optics EXPRESS

with

$$A = (2.696582\omega^5 + 2.030805\omega^4 - 3.172874\omega^3 + 3.790689\omega^2 - 5.725337\omega + 0.277823) - i(0.836455\omega^5 - 0.862450\omega^4 - 8.980376\omega^3 + 1.498696\omega^2).$$

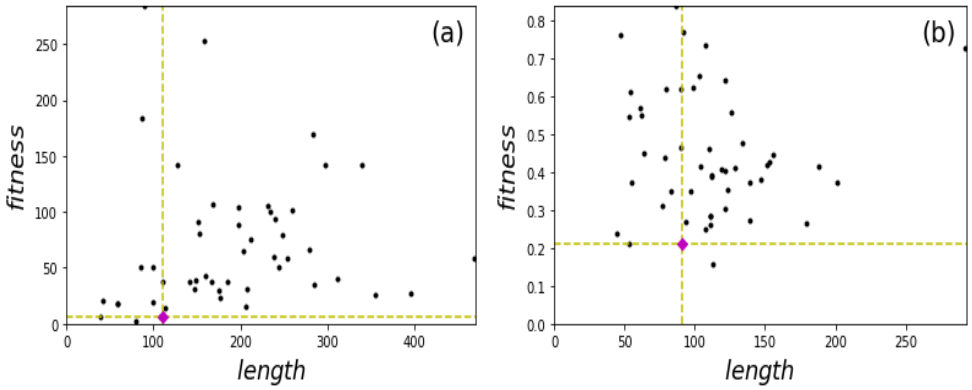


Fig. 6. Optimal solutions retrieved through the SR scheme from 50 different initial states. The pink diamonds represent the closed-form expression found considering the same initial state when the target data were generated: (a) with Eq. (12) and (b) with Eqs. (4) and (12).

The algebraic form of Eq. (13) could be considered again as a failure of the SR scheme. However, this result can explained by a reasoning similar to that used in the case of transparent dielectrics. That is, if we develop Eq. (12) we arrive to

$$\epsilon^T(\omega) = 3.01 + \frac{(-1.066422\omega^2 + 0.006645) - i\omega 0.013325}{(64.251919\omega^4 - 0.564148\omega^2 + 0.001) + i(0.876444\omega^3 - 0.002399\omega)}. \quad (14)$$

A visual inspection of Eqs. (13) and (14) suggests a behavior similar to that observed for lossless dielectrics. Putting aside the somehow expected differences between the terms in the fractions' numerator and denominator, the SR-scheme retrieves a complex closed-form expression whose algebraic form is similar to that of the target dielectric function. The predictions of Eq. (13) are shown in Fig. (7). For consistency we use the same markers as in the case of lossless dielectrics and the target data are depicted with blue dots, whereas the data predicted from ellipsometric and spectral data are respectively depicted with red upside-down and yellow triangles.

With reference to Fig. 7(a), the visual agreement between the target and predicted imaginary parts is confirmed by a $MARE[\epsilon_{i_e}^{SR}, \epsilon_i{}^T] = 0.482$. This behavior is also observed in Fig. 7(b), where the predicted and the target real parts are similar and they give a $MARE[\epsilon_{r_e}^{SR}, \epsilon_r{}^T] = 0.525$.

In Fig. 8(c), we compare the reflectance spectrum computed using the data predicted by Eq. (14) in Eq. (4). Excepting the slight differences between 600 and 700 nm, both spectra present a similar behavior.

With reference to Fig. 7(b), the green diamonds illustrate the result obtained when $\Re\{\epsilon_e^{KK}(\omega)\}$ is computed from $\Im\{\epsilon^{SR}(\omega)\}$ through K-K relations. The comparison between $\Re\{\epsilon_e^{KK}\}$ and $\Re\{\epsilon^T\}$ leads to $MARE[\Re\{\epsilon_e^{KK}\}, \Re\{\epsilon^T\}] = 2.588$ and this low error value suggests that dielectric function found is casual. Although not shown here for space reasons, the results of computing the imaginary part from the real part predicted by SR seem to confirm the causality of the dispersion model retrieved.

To close this section, we search for the dispersion model using spectral information as input to the SR-scheme. The results of 50 different realisations of the SR-scheme are shown in Fig. 6(b),

Optics EXPRESS

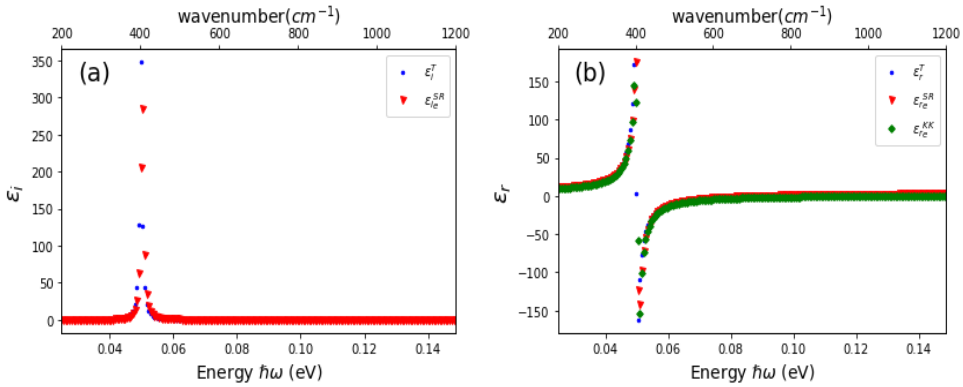
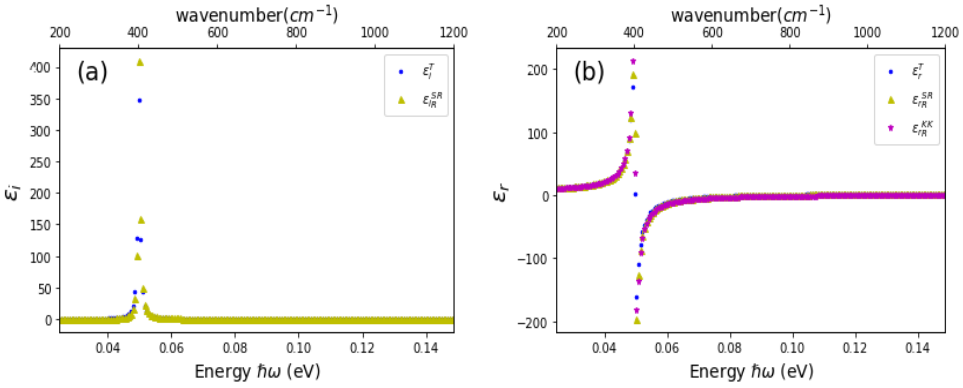


Fig. 7. (a) Imaginary part and (b) real part of *MgO* (blue dotted line) compared with the dielectric functions predicted by SR using ellipsometric data (red upside down triangles). The respective real parts in (b) computed through K-K relations are depicted with green diamonds.



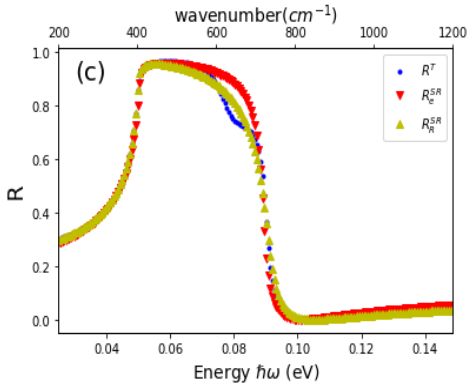


Fig. 8. (a) Imaginary part and (b) real part of *MgO* (blue dotted line) compared with the dielectric functions predicted by the expressions retrieved using reflectance data (yellow triangles). The respective real parts in (b) computed through K-K relations are depicted with purple stars. In (c) we compare target reflectance spectrum (blue dotted line) with the spectra predicted by SR using ellipsometric data (upside down triangles) and reflectance data (yellow triangles).

Optics EXPRESS

where the pink diamond corresponds to the same initial state considered when the search was done using ellipsometric data. As expected, there is a significant effect of the input data on the fitness value and the length of the solutions found. This can be also seen writing the closed-form expression related to the pink diamond that has the form

$$\epsilon_R^{SR}(\omega) = \frac{-0.287023\omega + 0.046929 + i(0.304372\omega^3 + 0.941903\omega^2 - 9.623235\omega + 0.865766)}{\omega^2 - 0.201880\omega + 0.008484 - i(3.238\omega^2 + 1.923933\omega - 0.103801)}. \tag{15}$$

The difference between the algebraic forms of Eqs. (12) and (15) is due, on the one hand, to the fact that the relationship between the real and imaginary parts of the dielectric function is lost because the squared modulus of the reflection coefficient is used. Also, when spectral information is taken as input to the SR, the problem is reduced to finding the best fit between the target and the predicted spectra regardless of the algebraic form of the dielectric function that generates the latter. The direct consequence of this situation is that the SR-scheme will find a complex function that, as illustrated in Figs. 8(a) and (b) accurately predicts the real and imaginary parts of the target dielectric function without resembling their algebraic forms. Furthermore, the reflectance spectrum generated with the expression retrieved agrees with the target one, as shown in Fig. 8(c).

The result obtained from computing $\Re\{\epsilon_R^{KK}(\omega)\}$ through K-K relations is depicted in Fig. 8(b) with purple stars and the related MARE value is 2.644. The agreement between $\Re\{\epsilon^T(\omega)\}$ and $\Re\{\epsilon_R^{KK}(\omega)\}$, together with the fairly low MARE, suggest that the SR scheme was able to find a causal model, although its algebraic form is not that of the target dielectric function. This situation could be useful for the resolution of certain problems where there is not an well established model.

4. Conclusions

In this contribution, which can be considered as a proof of concept, we illustrate how SR can be a suitable computational tool to model the optical properties of transparent and absorbing dielectric materials. The numerical evidence shows that, in addition to not requiring any assumptions about the algebraic form of the model being searched, this machine learning glass-box can treat complex functions, a possibility that to our knowledge had not been explored so much before. This fact led to find closed-form expressions that verify K-K relations. This is something that recent works making use of approaches based on AI, to describe the causal of behavior different materials, are not yet able to do. Thus, SR could complement the versatility of the approaches just mentioned.

Another finding of this work is that a closed expression with a complicated algebraic form does not necessarily imply a failure of the regression scheme or a violation of the parsimony principle. The results of our numerical experiments suggest that SR favors the minimization of the lost error (fitness function) rather than to keep the algebraic complexity low. That is, the regression scheme finds an equivalent but not simplified version of the mathematical model. While it is true that this fact could make difficult the physical interpretation of the expression found, the search for a compact expression depends on the problem studied. Thus, increasing systematically the parsimony pressure may limit the performance of SR and lead it to non-optimal solutions.

Although the results of this work are encouraging, further work is still required. The use of a multi-objective approach could contribute to simultaneously reduce the algebraic complexity of the expressions to be retrieved and the lost error. Furthermore, in this exploratory work we only considered reflectance spectra at normal incidence. Taking into account transmitance information, oblique incidence and the polarization could contribute to constrain the search space.

This contribution opens the possibility to use SR in problems where there are not well established models or the existing ones are not well suited. Moreover, there are not visible limitations to extend its application to more complex cases involving, for example, anisotropic materials or multilayer structures.

Optics EXPRESS

Disclosures. The authors declare no conflicts of interest.

Data availability. No data were generated or analyzed in the presented research.

References

1. A. Roger and D. Maystre, "Inverse scattering method in electromagnetic optics: Application to diffraction gratings," J. Opt. Soc. Am. **70**(12), 1483–1495 (1980).
2. S. Robert, A. M. Ravaud, S. Reynaud, S. Fourment, F. Carcenac, and P. Arguel, "Experimental characterization of subwavelength diffraction gratings by an inverse-scattering neural method," J. Opt. Soc. Am. A **19**(12), 2394–2402 (2002).
3. G. M. Sacha and P. Varona, "Artificial intelligence in nanotechnology," Nanotechnology **24**(45), 452002 (2013).
4. K. Yao, R. Unni, and Y. Zheng, "Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale," Nanophotonics **8**(3), 339–366 (2019).
5. T. Coen, H. Greener, M. Mrejen, L. Wolf, and H. Suchowski, "Deep learning based reconstruction of directional coupler geometry from electromagnetic near-field distribution," OSA Continuum **3**(8), 2222–2231 (2020).
6. A.-B. Djurišic, J.-M. Elazar, and A.-D. Rakic, "Modeling the optical constants of solids using genetic algorithms with parameter space size adjustment," Opt. Commun. **134**(1-6), 407–414 (1997).
7. A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski, "Optical properties of metallic films for vertical-cavity optoelectronic devices," Appl. Opt. **37**(22), 5271–5283 (1998).
8. A. Vial, A.-S. Grimault, D. Macías, D. Barchiesi, and M. L. de la Chapelle, "Improved analytical fit of gold dispersion: Application to the modeling of extinction spectra with a finite-difference time-domain method," Phys. Rev. B **71**(8), 085416 (2005).
9. K. S. Banerjee and I. Takayanagi, "Computing complex dispersive refractive indices from thin-film optical properties of materials," Proc. SPIE **11105**, 11105T (2019).
10. P. Vukusic and D. Stavenga, "Physical methods for investigating structural colours in biological systems," J. R. Soc. Interface. **6**(suppl_2), s133–s148 (2009).
11. S. Yoshioka and S. Kinoshita, "Direct determination of the refractive index of natural multilayer systems," Phys. Rev. E **83**(5), 051917 (2011).
12. D. Macías, A. Luna, D. Skigin, M. Inchaussandague, A. Vial, and D. Schinca, "Retrieval of relevant parameters of natural multilayer systems by means of bio-inspired optimization strategies," Appl. Opt. **52**(11), 2511 (2013).
13. D. Skigin, M. Inchaussandague, D. Macías, and A. Vial, "Determination of the spectral-dependent refractive index of a single layer in a natural multilayer system: comparison of different approaches," Appl. Opt. **56**(7), 1807–1816 (2017).
14. O. Khatib, S. Ren, J. Malof, and W. J. Padilla, "Learning the physics of all-dielectric metamaterials with deep lorentz neural networks," Adv. Opt. Mater. **10**(13), 2200097 (2022).
15. A.-P. Blanchard-Dionne and O. J. F. Martin, "Teaching optics to a machine learning network," Opt. Lett. **45**(10), 2922–2925 (2020).
16. M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," Science **324**(5923), 81–85 (2009).
17. M. Quade, M. Abel, K. Shafi, R. K. Niven, and B. R. Noack, "Prediction of dynamical systems by symbolic regression," Phys. Rev. E **94**(1), 012214 (2016).
18. A. Radi, "Prediction of nonlinear system in optics using genetic programming," Int. J. Mod. Phys. C **18**(03), 369–374 (2007).
19. S. K. H. Amr Radi, "Genetic programming for modeling a nonlinear optical system," Egypt. J. Solids **31**, 269–275 (2008).
20. Y. Wang, N. Wagner, and J. M. Rondinelli, "Symbolic regression in materials science," MRS Commun. **9**(3), 793–805 (2019).
21. M. T. Silviu-Marian Udrescu, "Ai feynman: A physics-inspired method for symbolic regression," Sci. Adv. **6**(16), eaay2631 (2020).
22. R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics; New Millennium Ed.* (Basic Books, 2010). Originally published 1963–1965.
23. J. R. Koza, *Genetic Programming - On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)* (MIT, 1993).
24. T. McConaghy, "FFX: Fast, scalable, deterministic symbolic regression technology," in *Genetic Programming Theory and Practice IX*, R. Riolo, E. Vladislavleva, and J. H. Moore, eds. (Springer, 2011), Chap. 13, pp. 235–260.
25. R. Poli, W. B. Langdon, and N. F. McPhee, *A Field Guide to Genetic Programming* (lulu.com, 2008).
26. S. Luke and L. Panait, "A comparison of bloat control methods for genetic programming," Evol. Comput. **14**(3), 309–344 (2006).
27. W. B. Vinícius Veloso de Melo, "Automatic feature engineering for regression models with machine learning: an evolutionary computation and statistics hybrid," Inf. Sci. **430-431**, 287–313 (2018).
28. "Genetic programming in python, with a scikit-learn inspired API: gplearn," https://gplearn.readthedocs.io/en/stable/index.html.
29. J. McCarthy, "Recursive functions of symbolic expressions and their computation by machine, part i," Commun. ACM **3**(4), 184–195 (1960).

Optics EXPRESS

30. "Sympy's documentation https://docs.sympy.org/latest/index.html,".
31. C. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, 1983), Chap. 2, pp. 30–33.
32. M. Heidarian, M. Karimnezhad, M. Schaffie, and M. Ranjbar, "A new empirical correlation for estimating bubble point pressure using the genetic algorithm," Geol. Geophys. Environ. **43**(1), 33 (2017).
33. I. H. Malitson, "Interspecimen comparison of the refractive index of fused silica," J. Opt. Soc. Am. **55**(10), 1205–1208 (1965).
34. C. Z. Tan, "Determination of refractive index of silica glass for infrared wavelengths by ir spectroscopy," J. Non-Cryst. Solids **223**(1-2), 158–163 (1998).
35. R. Tilley, *Colours Due to Refraction and Dispersion* (John Wiley and Sons, Ltd., 2010), Chap. 2, pp. 49–90.
36. K. Ohta and H. Ishida, "Comparison among several numerical integration methods for kramers-kronig transformation," Appl. Spectrosc. **42**(6), 952–957 (1988).