T. N. Kupenova [*]

AN INDUCTIVE ALGORITHM
FOR SMOOTH APPROXIMATION OF FUNCTIONS

[*] Permanent address: Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Tzarigradsko chaussee 72, 1784 Sofia, Bulgaria; e-mail: kupenova@inrne.bas.bg

Купенова Т. Н.   Е11-2011-97

Индуктивный алгоритм гладкой аппроксимации функций

Предлагается индуктивный алгоритм гладкой аппроксимации функций, основанный на методе регуляризации Тихонова для конкретной формы функционала Тихонова. Параметр регуляризации находится по принципу невязки. Принцип эвристической самоорганизации используется для оценки некоторых параметров аппроксимирующей функции.

Kupenova T. N.   Е11-2011-97

An Inductive Algorithm for Smooth Approximation of Functions

An inductive algorithm is presented for smooth approximation of functions, based on the Tikhonov regularization method and applied to a specific kind of the Tikhonov parametric functional. The discrepancy principle is used for estimation of the regularization parameter. The principle of heuristic self-organization is applied for assessment of some parameters of the approximating function.

INTRODUCTION

In spite of a big amount of existing methods for smooth approximation and software developed, they are frequently insufficient to meet the need of the oncoming recent technological developments. For instance, certain difficulties exist in the case of many dimension, large data volumes ($\sim 100\,000$ data points are usual for terrain modeling); nonhomogeneous data (clusterized data), finding the derivatives with high accuracy.

Also, some methods need the value of some parameter to be set up ad hoc, manually, so it is necessary to find methods for automatic choice of its best value. The subject of the present paper is to introduce an inductive algorithm for smooth approximation of functions, based on the Tikhonov regularization and applied to a specific kind of the Tikhonov parametric functional.

The discrepancy principle is used for estimation of the regularization parameter α.

The principle of heuristic self-organization is applied for assessment of some parameters and to increase the predictive ability of the model. The efficiency of the algorithm is illustrated using two examples. The first one is approximation of simple model function, given in N points on the interval $[a, b]$. Artificial normally distributied error was added using the random numbers generator. The second one is estimation of the gain and the contrast gain of the d-wave in the electroretinogram.

APPROXIMATION PROBLEM

The approximation problem can be posed in the following way:
Given:
$$Az = u \equiv (z(x_1), z(x_2), \ldots, z(x_N)), z \in Z, \tag{1}$$
where $\{x_i, i - 1, \ldots, N\}$ is a lattice of knots on $[a, b]$, Z is metric space.
Find the element:

$$z \in Z: \quad Az = u, \quad u \in R_N.$$

Let u_σ and $\sigma > 0$ are given, instead of u, so that $\rho(u_\sigma, u) < \sigma$, where ρ is the metric in R_N, i.e. the data include experimental error. This problem is ill-posed because of the existence of an infinite number of solutions (broken condition

of uniqueness). In addition, A^{-1} could not be continuous (broken condition of stability).

To solve this problem, one must introduce some additional information about the solution such as assumption on the smoothness. Following the Tikhonov regularization method [1], the solution is the function $g(x) \in Z$ that minimizes the Tikhonov parametric functional:

$$
\begin{aligned}
M^\alpha(z, u_\sigma) &= \rho^2(Az, u_\sigma) + \alpha\Omega(z), \\
M^\alpha(g, u_\sigma) &= \inf_{z \in Z} M^\alpha(z, u_\sigma) = \inf_{z \in Z} \left(\rho^2(Az, u_\sigma) + \alpha\Omega(z) \right).
\end{aligned}
\tag{2}
$$

Parameter of regularization α is estimated from the condition for the discrepancy:

$$
\rho(Ag, u_\sigma) = \sigma.
\tag{3}
$$

The regularization method is determined by the choice of:
- solution space Z and stabilizer $\Omega(z)$;
- estimation of the parameter of regularization α according to additional information (experimental error σ).

A large class of approximation methods including the splines are based on the condition of smoothness. It can be formulated within the variational principle as a minimization of some smoothness functional. For example, $\Omega(z) = \int\limits_a^b \left| \dfrac{d^2 g}{dx^2} \right|^2 dx$ is the case of cubic smoothing spline. Probably the most general form of this approach was introduced by Talmi and Gilat [2].

BRIEF DESCRIPTION OF THE METHOD

We start with the definition of the «smoothness co-functional» (inner product) which is the central point of the theory.

Let $g(x)$ and $h(x)$ be elements of the space W of the complex analytic functions, defined and finite (they and all their derivatives) on $[a, b]$.

Then we can express this inner product in the form

$$
I(g, h) = \sum_{n=0}^{\infty} B_n \int\limits_a^b g(x)^{(n)*} h(x)^{(n)} dx,
\tag{4}
$$

where B_n are (arbitrary) nonnegative constants, chosen to ensure the convergence of the series.

For $g(x) \equiv h(x)$, this functional naturally induces semi-norm:

$$
I(g, g) = I(g) = \sum_{n=0}^{\infty} B_n \int\limits_a^b \left| g^{(n)}(x) \right|^2 dx,
\tag{5}
$$

2

called «smooth norm», because it represents in some sense the smoothness of $g(x)$.

Let $\Omega(g) = I(g)$.

Now the approximation problem can be formulated as follows:

Construct a series $g(x) = \sum\limits_{k=1}^{\infty} A_k \Phi_k(x)$, where $\{\Phi_k(x)\}$ is a set of «orthogonal» functions, i.e., $I(\Phi_k, \Phi_l) = \delta_{kl} I(\Phi_k)$, so that

$$
M^{\alpha}(z, F) = \rho^2(Ag, F) + \alpha\Omega(g) =
$$
$$
= \sum_{i=1}^{N} |F(x_i) - g(x_i)|^2 \, \omega_i + \alpha I(g) = \min, \quad (6)
$$

$F(x_i), i = 1, 2, \ldots, N$, are the experimental data, ω_i are weights, α is the parameter of smoothing (regularization), and $g(x)$ is the approximating function.

It is proved [2] that there exists one and only one solution:

$$
g(x) = \sum_{n=1}^{N} \lambda_j R(x, x_j), \quad (7)
$$

where

$$
R(x, y) = \sum_{k=1}^{\infty} \frac{\Phi_k(x)\Phi_k^*(y)}{I(\Phi_k)},
$$

and λ_j are found by the solution of the set of linear equations:

$$
g(x_i) = \sum_{j=1}^{N} \lambda_j \left[R(x_j, x_j) + \frac{\alpha}{\omega_j}\delta_{ij} \right] = F(x_i) \quad i = 1, \ldots, N. \quad (8)
$$

It is naturally to choose the weights ω_i in accordance with data errors $\sigma(x_i)$: $\omega_i = [\sigma(x_i)]^{-2}$.

The choice of smoothing parameter α is very important. Too small α leads to spiky curve. Too large α leads to smooth curve, but far from the data. The optimal value of α, such that the solution $g(x)$ is regularized, can be found from the discrepancy condition (3).

Among the advantages of this methods are: its ability to obtain a good approximation, not only of the data points, but also of their first and higher derivatives; ability to include additional information and constraints given by the theory or experiment. The method is extended to spaces of arbitrary dimension:

$$
I(g) = \sum_{n_1, n_2, \ldots, n_L}^{\infty} B_{n_1 n_2 \ldots n_L} \int dV \left| \frac{\partial^{n_1 + n_2 + \ldots + n_L} g(x_1, x_2, \ldots, x_L)}{\partial x_1^{n_1} \partial x_2^{n_2} \ldots \partial x_L^{n_L}} \right|^2.
$$

3

PRACTICAL REALIZATION

There is some freedom in choice of coefficients B_n. This allows one to express $R(x, y)$ by simple functions according to the choice of B_n [2]. To distinguish between «near» and «distant» points, a measure D, called «correlation length», is introduced. D can be chosen somewhat arbitrarily and it is related to the average density of points. The choice of D will be discussed later.

In this paper we choose $B_n = \dfrac{D^{2n}}{(2n)!}$. Then $R(x, y)$ has the following form:

$$R(x, y) = 1/(2\, D \cosh{(\pi r/2D)}).$$

Fortran codes for smooth approximation for functions of one, two, three dimensions have been developed based on this method. They were applied to many physical problems [3–6].

CHOICE OF CORRELATION LENGTH D

To find the optimal value of D, we use an inductive approach based on the principle of heuristic self-organization [7] and principle of external complement [8]. According to these principles, an optimal mathematical model corresponds to the minimum of some «external» criterion, i.e., based on additional fresh information, which was not used for model construction. To do this, we perform the following steps:

1. Split data sample into two subsamples:
- learning set A: N_L data points, distributed as uniformly as possible, and
- the rest ($N_V = N - N_L$ data points) — validation set B.

2. Estimate coefficients $\lambda_j, j = 1, \ldots, N$ of the approximating function g and the smoothing parameter α using the proposed method (Eqs. $(3), (6)$ and (8)) and learning set A.

3. Calculate values of the external criterion for the approximating function g using the validation set B for different values of the correlation length D.

4. Choice of the best model (best D) corresponding to the minimal value of external criterion.

We have to achieve two aims:

- For the first aim — to obtain an optimal estimation of the parameters of the approximation $\{\lambda_j\}$ and smoothing parameter α, we use the criterion for minimum of Tikhonov functional (6) over some subset of data (training set A).

- For the second aim — finding the model with good prognostic properties (optimal D), we need second criterion over different subset of data (validation set B). In this paper we use the criterion for minimum of maximum deviation on the validation set B.

The advantages of this approach are:

• Better level of automatization due to automatic choice of the best value of the correlation length D. One needs to enter initial data sample and type of external criterion only.

• Improve the predictive ability of the model. In particular, we find the approximating function by maximizing its performance on some set of data. However, its performance is determined not only by its accuracy on this set of data, but also by its ability to predict well on unseen data.

There are also some problems.

• Only one part of the data is used for fitting. This leads to higher variance.

• An open question is how big should the learning set be? Relatively large learning sets lead to overfitting, relatively small training sets lead to under-fitting.

APPLICATIONS OF THE METHOD

The efficiency of the method is illustrated by two examples. The first one is the approximation of simple model function $F(x)$. The second one is estimation of the gain and the contrast gain of the d-wave in the electroretinogram.

Approximation of Model Function. The function $F(x) = \dfrac{1}{1+x^2}$ is given in N points on the interval $[a, b]$. An artificially generated normally distributed random errors are added to analytically computed values with standard deviation σ:

$$F_\sigma(x_i) = F(x_i) + \sigma_i.$$

The noised values $F_\sigma(x_i)$ are applied as input for checking the algorithm. We divide given data into two sets in the following way: the points with odd i form the learning set, the rest — validation set. The coefficients $\lambda_j, j = 1, \ldots, N$, are determined for different values of D using the proposed method (Eqs. (3), (6) and (8)) and the data from learning set. To find the optimal value of the parameter D, we use the data from the validation set and the criterion for minimum of maximum error Δ over the validation set:

$$\Delta = \max_{i=1,N_V} |g(x_i^*) - F_\sigma(x_i^*)| = \min, \tag{9}$$

where $g(x_i^*)$ and $F(x_i^*)$ are the approximating function and «experimental value», and N_V is the number of points in the validation set.

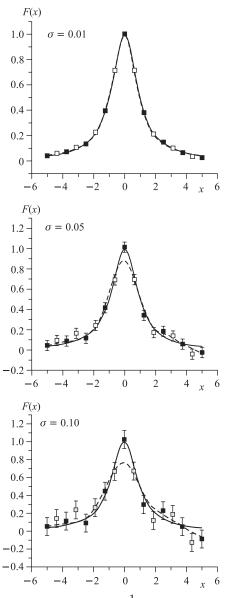The results for $N = 17$ and $\sigma = 0.01$, $\sigma = 0.05$, $\sigma = 0.1$ are shown in Fig. 1.

5

Fig. 1. Approximation of function $F(x) = \dfrac{1}{1 + x^2}$, $N = 17$, $x \in [-5, 5]$. Analitically computed function F — solid lines, approximating function F_{appr} — dashed lines, «experimental data» F_σ — symbols

ESTIMATION OF THE GAIN AND THE CONTRAST GAIN OF THE D-WAVE IN THE ELECTRORETINOGRAM

Electroretinography measures the electrical responses of various cell types in the retina. Control of the gain (sensitivity) and the contrast gain (contrast sensitivity) of the visual responses is very important for the visual system. It allows discrimination between objects of different intensity and contrast in conditions of ambient illumination, varying in an enormously wide range (10^{11}), while using a restricted dynamic range of the responses of the visual cells.

The gain and the contrast gain can be estimated on the basis of the measured response vs. intensity function $V(I)$:

- $V(I)$ describes the response vs. intensity function of the cells in the peripheral retina.
- $dV/d(I)$ is used for estimation of the gain (sensitivity) of the light responses.
- $dV/d(\log I)$ is a measure of contrast gain (contrast sensitivity) of these responses.

A mass electrical response from the retina (electroretinogram — ERG) is easily recorded and widely used in the research or clinical practice for evaluation of functional state of the retina as well as an objective criterion for the sensitivity of the visual system, because there is a good correlation between ERG and psychophysical sensitivity.

The proposed method was applied for estimation of the gain and the contrast gain. A computer code is developed for estimation of the gain and the contrast gain according to the experimentally measured values of $V(I)$. We split data set into two subsets in the following way: the points with odd i form the learning set, the rest — validation set. The coefficients $\lambda_j, j = 1, \ldots, N$, are determined for different values of D using the proposed method (Eqs. (3), (6) and (8)) and the data from the learning set. To find the optimal value of the parameter D, we use the data from the validation set and the criterion for minimum of maximum error Δ over the validation set.

The results of application of the algorihm for estimation of the gain and the contrast gain of the d-wave in the electroretinogram recorded from turtle retina are shown in Fig. 2. The curves calculated for different values of the parameter D are drawn by different kind of lines. It is seen that the best results for $V(\log(I))$, as well as for gain and the contrast gain, are for the value of D corresponding to minimum of external criterion Δ.

CONCLUSION

We have presented an efficient algorithm for smooth approximation of data. The main advantages of this approach are:

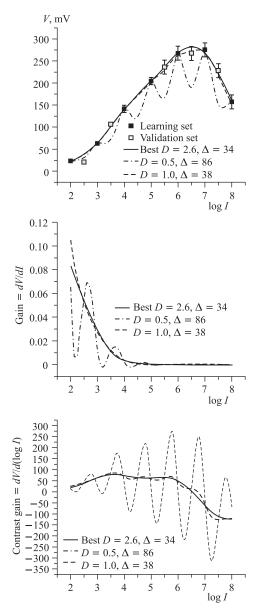1. The Tikhonov regularization ensures stability of the solution.

Fig. 2. Estimation of the gain and the contrast gain of the d-wave in the electroretinogram recorded from turtle retina

2. The specific choice of smoothing functional has many advantages:

• Simultaneously good approximations for the function as well as for its derivatives;

• high degree of flexibility — it is possible to incorporate into it additional information such as the derivatives or some constraints from the theory and experiment;

• a smooth connection of data points belonging to different ranges of a given parameter;

• it is easy to extend the method for arbitrary dimension.

3. The inductive approach based on the principle of heuristic self-organization ensures better level of automatization and improves the predictive ability of the model.

In conclusion, we believe that the method presented here will be useful for a large variety of problems encountered in the field of science and technology, where here is a need to fit data by analytical functions.

REFERENCES

1. *Tikhonov A. N., Arsenin V. Ya.* Methods of Solving Ill-Posed Problems. M.: Nauka, 1979 (in Russian).

2. *Talmi A., Gilat G.* Method for Smooth Approximation of Data // J. Comp. Phys. 1977. V. 23. P. 93–123.

3. *Stoyanova V., Kashchiev D., Kupenova T.* Freezing of Water Droplets Seeded with Atmospheric Aerosols and the Ice Nucleation Activity of the Aerosols // J. Aerosol Sci. 1994. V. 25, No. 5. P. 867–877.

4. *Kupenova T., Kupenova P.* Method for Estimation of the Gain and the Contrast Gain of the D-Wave in the Electroretinogram // Proc. of the 5th General Conference of the Balkan Physical Union BPU-5 / Eds. S. Jokic, I. Milosevic, A. Balaz, Z. Nikolic, Serbian Physical Society, Belgrade, 2003. P. 1475–1478; http://www.phy.bg.ac.yu/jdf/bpu5/proceedings/Papers/SP10007.pdf

5. *Stoyanova V., Kupenova T.* Laboratory Analysis and Ice Forming Characterization of Fine Dust Particles from the Lowermost Atmosphere of Sofia // Compt. Rend. Acad. Bulg. Sci. 2007. V. 60, No. 5. P. 545–550.

6. *Kupenova P., Popova E., Vitanova L.* Gabaa and Gabac Receptor Mediated Influences on the Intensity-Response Function of the b- and d-Wave in the Frog Erg // Vision Res. 2008. V. 48. P. 882–892.

7. *Ivakhnenko A. G.* Heuristic Self-Organization in Problems of Engineering Cybernetics // Automatica, 1970. V. 6.

8. *Beer S.* Cybernetics and Management. London: The English Universities Press LTD, 1967.