The Steepest Descent Method Using the Empirical Mode Gradient Decomposition

Vasiliy Esaulov and Roman Sinetsky
Platov South-Russian State Technical University, Prosvescheniya Str. 132, Novocherkassk, Russia
rmsin@srspu.ru, esaul_va@mail.ru

Keywords: Steepest Descent Method, Gradient Descent Method, Empirical Mode Decomposition, Optimization Problem.

Abstract: The aim of the article is to study the possibility of improving gradient optimization methods. The leading approach to the chosen concept is based on the possibility of a featured description of the gradient that sets the direction of the search for a solution. A modification of the method of steepest descent of global optimization based on the Hilbert-Huang transform is proposed. The proposed solution is based on the decomposition of the gradient of the objective function into empirical modes. The main results of the work are iterative optimization methods, in which, in addition to the gradient, its empirical modes are also taken into account. New estimates of the descent step are obtained, which could not be deduced in the classical formulation of the steepest descent method. Their correctness is due to the fact that in the absence of the possibility of gradient decomposition, they are reduced to existing estimates for the steepest descent method. The theoretical significance of the results lies in the possibility of expanding the existing gradient methods by a previously not used gradient description method. The practical significance is that the proposed recommendations can help accelerate the convergence of gradient methods and improve the accuracy of their results. Using the Python language, computational experiments were carried out, as a result of which the adequacy of the proposed method and its robustness were confirmed.

1 INTRODUCTION

The optimization problem is a significant mathematical model in a wide class of disciplines. Its methods are applied in areas such as computer-aided design, machine learning, mathematical modeling, and others. As one of the main statements of the optimization problem, we will further consider the problem of finding the minimum of a function. Let the task of finding the minimum

$$F(X) \to \min, \ X \in R^n, \tag{1}$$

where $F(X)$ – objective function; X – objective function parameters.

The formula for the coordinate descent process for (1) in the case of applying the gradient has the form

$$X_{k+1} = X_k - \lambda_k \nabla F(X_k), \ k = 0,1,2...., \tag{2}$$

where $\nabla F(X_k)$ – objective function gradient; X_k, X_{k+1} – objective function parameters at k and k+1 iteration respectively; λ_k – step value, $\lambda_k \geq 0$.

The essence of the steepest descent method is the selection of such λ_k, where, with a known X_k, the condition is satisfied.

$$F(X_k - \lambda_k \nabla F(X_k)) \to \min, \ \lambda_k \geq 0. \tag{3}$$

Let us consider the possibility of modifying the steepest descent method based on the representation of the gradient of the objective function in some basis.

Indeed, we can consider gradient $\nabla F(X_k)$ as a discrete one-dimensional signal having a length equal to the dimension of the search space. This makes it possible to apply the methods of signal processing theory to it.

Having a spatial decomposition of the gradient of the objective function in some basis, one can both improve the convergence of gradient methods and get the opportunity to synthesize their modifications with fundamentally new properties.

2 EMPIRICAL MODE DECOMPOSITION OF GRADIENT

Let us consider the possibility of applying the empirical mode decomposition method to the gradient of the objective function [2, 3]. The principle of decomposition into empirical modes developed relatively recently. Its main specialization is the analysis of non-stationary processes. It is quite well established in a broad range of problems [6, 7].

One of the significant advantages of the empirical mode decomposition (EMD) method is that it does not require a choice of basis. Unlike Fourier or wavelet analysis, a mathematical apparatus is less developed for it. However, this fact does not reduce interest in studying the effectiveness of its application for practical problems.

Let us consider some general features of the empirical mode method. The basic functions used in the decomposition are extracted directly from the original signal. This, in turn, allows us to take into account its individual structural features.

The qualitative basis of the apparatus of empirical modes is to use the multiple addition of white noise to the signal. Next, the average value of the distinguished components is calculated by the classical method of decomposition as the result.

As a result of decomposition, the signal is presented in the time-frequency domain, which allows revealing hidden modulations and energy concentration regions. Since the decomposition is based on the data of a specific local time domain of the signals, it is also applicable to non-stationary signals. Using EMD, it is possible to determine the instantaneous frequency as a function of time, which allows you to get a clear idea of the internal structure of the signal [2–5].

An *empirical mode* (or *intrinsic mode function*, IMF) is such a function that has the following properties [7]:

1) The number of function extrema (maxima and minima) and the number of zero intersections should not differ by more than one.
2) At any point, the average value of the envelopes defined by local maxima and local minima should be zero.

IMF is an oscillatory function, but instead of a constant amplitude and frequency, as in a simple harmonic, IMF can have a variable amplitude and frequency, as functions of an independent variable (time, coordinate, etc.).

The first property guarantees that the local maxima of the function are always positive, the local minima are respectively negative, and between them, there always are intersections of the zero line.

The second property ensures that the instantaneous frequencies of the function will not have undesirable fluctuations resulting from the asymmetric waveform.

Any function and any arbitrary signal that initially contains an arbitrary sequence of local extrema (minimum 2) can be divided into the IMFs family and the residual trend. If the data are devoid of extrema, but contain inflection points ("hidden" extrema of superimposing mode functions and steep trends), then signal differentiation can be used to "open" extrema [8, 9].

Suppose that there is an arbitrary signal $x(t)$. The essence of the EMD method consists in sequentially calculating the functions of the empirical modes $c_j(t)$ and the residues $r_j(t) = r_{j-1}(t) - c_j(t)$, where $j = 1, 2, ..., n$ at $r_0 = x(t)$. The decomposition result will be the representation of the signal as a sum of mode functions (IMFs) and the final residual [6–9]:

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t)\,, \tag{4}$$

where n is the number of IMFs that is established during the calculations.

3 EMD ALGORITHM

The block diagram of the EMD algorithm is presented in Figure 1 [4–6].

The EMD algorithm consists of the following operations:

1) For any data $x(t)$, all local extrema are identified.
2) Based on the extrema, the upper, $u(t)$, and lower, $l(t)$, envelopes are formed (in this case, cubic spline interpolation can be used).
3) Envelope mean value is calculated as
4) $m(t) = [u(t) + l(t)] / 2$.
5) The difference between the original signal and the average value is considered as IMF
6) $h(t) = x(t) - m(t)$.
7) The current $h(t)$ value is evaluated for IMF compliance.
8) If $h(t)$ does not satisfy the definition of IMF, go to steps 1-5. Otherwise, the IMF is accepted as component $c(t)$.
9) The residual function $r(t) = x(t) - c(t)$ is determined. Steps 1 to 6 are repeated for $r(t)$.
10) The operation ends when $r(t)$ contains no more than one extremum.

So, using the EMD method, let decomposition of $\nabla F(X_k)$ constructed on basis of modes $\{H_i(X_k)\}$, i=1,…,m in such a way that

$$\nabla F(X_k) = \sum_{j=1}^{m} H_j(X_k). \quad (5)$$

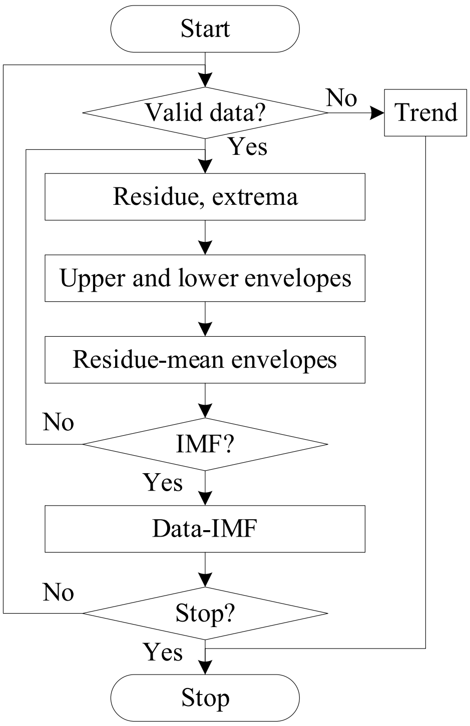

Figure 1: The block diagram of the EMD algorithm.

Set a descent for each of the mods

$$X_{k+1}^{j} = X_k - \lambda_k H_j(X_k),\ j = 1,2,\ldots,m. \quad (6)$$

The obtained set of points $\{X^i_{k+1}\}$, i=1,…,m can be considered as a combination of some alternative results obtained by approximating the gradient of the objective function while maintaining its structural properties. At a qualitative level, this makes the process of finding the optimum nonlocal.

The presence of many possible alternatives to the solution can accelerate the convergence of the search process. On the other hand, it can be expected that it will be more stable in situations where the initial approximation is given far enough from the optimal solution.

4 DESCENT STEP SELECTION

From statements (5), (6), two variants of modification of the search rule for step λ_k were formulated in the paper. In the first case, we can require the following condition

$$\sum_{j=1}^{m} F(X_{k+1}^{j}) = \sum_{j=1}^{m} F(X_k - \lambda_k H_j(X_k)) \to \min. \quad (7)$$

Thus, the above condition consists in finding a step that minimizes the sum of the function values calculated at points obtained by descent along all components of the EMD of gradient $\nabla F(X_k)$. The decrease in the value of the function on average over the totality of values on the set $\{X^i_{k+1}\}$, i=1,…,m allows us to talk about the global nature of optimization.

We can estimate the optimal descent step in (7). Expand the left side in a Maclaurin series

$$\begin{aligned} F(X_{k+1}^{j}) &\approx F(X_k) - \lambda_k \left(\nabla F(X_k), H_j(X_k)\right) + \\ &+ \lambda_k^2 H_j^T(X_k) G(X_k) H_j(X_k) \end{aligned}, \quad (8)$$

where $G(X_k)$ – Hessian matrix of objective function;

Then for (7) we obtain the following approximation

$$\begin{aligned} f(\lambda_k) &= \sum_{j=1}^{m} F(X_{k+1}^{j}) \approx \\ &\approx \sum_{j=1}^{m} \Big(F(X_k) - \lambda_k \left(\nabla F(X_k), H_j(X_k)\right) + \\ &+ \lambda_k^2 H_j^T(X_k) G(X_k) H_j(X_k) \Big) \end{aligned}$$

The necessary optimality condition in this case has the form

$$\frac{df(\lambda_k)}{d\lambda_k} = 0. \quad (9)$$

From (9) we obtain an estimation for the step

$$\begin{aligned} \lambda_k &= \frac{\sum_{j=1}^{m} \left(\nabla F(X_k), H_j(X_k)\right)}{\sum_{j=1}^{m} H_j^T(X_k) G(X_k) H_j(X_k)} = \\ &= \frac{\left(\nabla F(X_k), \nabla F(X_k)\right)}{\sum_{j=1}^{m} H_j^T(X_k) G(X_k) H_j(X_k)} \end{aligned} \quad (10)$$

Expression (9) allows us to estimate the descent step taking into account several empirical modes, which are the levels of gradient decomposition in the space of empirical modes.

The solution to problem (7) can be used in two ways. The first of them is that the obtained λ_k can be used to go over to the next approximation in (2). Another way is to iterate over alternatives from the set $\{X^i_{k+1}\}$, i=1,…,m. The point X^i_{k+1} at which the smallest value is reached can be considered as the next approximation to which process (7) can be reapplied.

Another search option for λ_k is to minimize the expression

$$\inf_{X \in \left\{X_{k+1}^i\right\}_{i=1}^m} F(X) \to \min . \quad (11)$$

It can be seen from the above that the use of EMD allows us to obtain many alternative solutions to problem (1). Moreover, it can be reformulated in terms of alternative expressions (7), (10).

5 EXPERIMENTS

As a technique that allows us to evaluate the algorithms proposed in (7), (9), we used the solution of test problems of multidimensional optimization problems reduced to statement (1). The optimization results were compared with those obtained using the standard steepest descent algorithm. The maximum number of iterations was set as 50000, the convergence error was set as 10^{-11}. As the first test function, a quadratic function of the form

$$F(X) = \sum_{i=1}^{n} i \cdot x_i^2 \quad (12)$$

was used.

The test value of the function is $F^* = 0$. The initial approximation has the form x_0 = (-2, 2, -2, 2, -2, 2, -2, 2, -2, 2,-2, 2,). The simulation results are shown in table 1.

As the second test function, the Rastrigin function was used [1]. The test value of the function is $F^* = 0$. The initial approximation has the form x_0 = (-5, 5, -5, 5, -5, 5, -5, 5, -5, 5,-5, 5,). The simulation results are shown in table 2.

As the third test function, the Rosenbrock function [1] was used. The test value of the function is $F^* = 0$. The initial approximation has the form x_0 = (-2, 2, -2, 2, -2, 2, -2, 2, -2, 2,-2, 2,). The simulation results are shown in table 1.

Table 1: Experiments results for function (12).

Method	Function value	Number of iterations
Method (2), (3)	0	108
Method (7) with search	0	103
Method (7), (2)	–	–
Method (11) with search	0	107

Table 2: Experiments results for Rastrigin function.

Method	Function value	Number of iterations
Method (2), (3)	0	14
Method (7) with search	0	12
Method (7), (2)	0	15
Method (11) with search	0	10

Table 3: Experiments results for Rosenbrock function.

Method	Function value	Number of iterations
Method (2), (3)	$4.485 \cdot 10^{-17}$	30905
Method (7) with search	$1.415 \cdot 10^{-16}$	30884
Method (7), (2)	$4.968 \cdot 10^{-17}$	10052
Method (11) with search	$9.4 \cdot 10^{-4}$	1924

6 CONCLUSION

From the presented results it is seen that the gradient decomposition in the case of applying the EMD method gives adequate optimization results. At the same time, when trying to combine it with the traditional steepest descent method, a situation of solution divergence may arise. On the other hand, the application of methods (7), (11) can lead to a decrease in the number of iterations in comparison with the traditional method of steepest descent. Thus, the possibilities of a refined search for the descent step that exist in (7), (11), as well as the choice of an approximation obtained from many alternative options, are the strengths of the method proposed in the work.

REFERENCES

[1] A.A. Zhiglyavskii, A.G. Zhilinskas, “Metody poiska global'nogo optimuma” [Methods of search of global optimum]. Moscow, Nauka Publ., 1991, p. 247.

[2] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung and H.H Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”, Proc. R. Soc. London Ser. A., vol. 454, 1998, p. 903.

[3] N.E. Huang, Z. Shen and S.R. Long, “A new view of nonlinear water waves”, the Hilbert spectrum, Annu. Rev. Fluid Mech, vol. 31, 1999, p. 417.

[4] N.E. Huang, “The Hilbert–Huang transform and its applications”, S.S.P. Shen. Singapore, World Scientific, 2005.

[5] K.T. Coughlin and K.K. Tung, “11-year solar cycle in the stratosphere extracted by the empirical mode decomposition method”, Adv. Space Res., vol. 34, 2004, p. 39.

[6] E.P.S. Neto, M.A. Custaud, C.J. Cejka, P. Abry, J. Frutoso, C. Gharib and P. Flandrin, “Assessment of cardiovascular autonomic control by the empirical mode decomposition”, Method. Inform. Med., vol. 43, 2004, p. 60.

[7] Z. Wu and N.E. Huang, “A study of the characteristics of white noise using the empirical mode decomposition method”, Proc. R. Soc. London, Ser. A., vol. 460, 2004, p. 1597.

[8] P.O. Pavlovichev and A.L. Priorov, “The empirical mode decomposition for reduction in speech signals”,

DSPA: Digital Signal Processing Issues, vol. 6, no. 2, 2016, pp. 398-403.

[9] A.K. Alimuradov, Y.S. Kvitka, A.P. Zaretskiy and A.P. Kuleshov, “Noiseproof processing of speech signals based on the complementary ensemble emperical mode decomposition”, Proceedings of Moscow Institute of Physics and Technology, vol. 3 (31), 2016, pp. 43-53.