ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

INFORMACION TECHNOLOGY. AUTOMATION

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ. АВТОМАТИЗАЦІЯ

UDC 517.9/519.6

A. Smorodin
Odessa National Polytechnic University, 1 Shevchenko Ave., Odessa, Ukraine, 65044; e-mail: Andrey.v.smorodin@gmail.com

PREDICTIVE CONTROL METHODS IN TASKS OF SEARCHING SADDLE POINTS

А.В. Смородин. **Методи прогнозування управління в задачах пошуку сідлових точок.** У статті представлені нові методи пошуку стаціонарних точок функції багатьох змінних, в тому числі сідлових. Такі завдання зустрічаються в різних галузях теоретичної і практичної науки, наприклад, у побудові сідлових точок в дизайні лінз, машинному або глибокому навчанні завдань опуклою оптимізації та нелінійного програмування (необхідні і достатні умови вирішення формулюються за допомогою сідлових точок функції Лагранжа і доводяться в теоремі Куна-Таккера. При навчанні нейронних мереж доводиться повторювати процес навчання на великих кластерах і перевіряти здатність до навчання мережі при різних функціях втрати і різній глибині мережі, тобто проводити тисячі запусків нових обчислень, де кожен раз оптимізується функція втрати на великих обсягах даних, тому будь-яке прискорення процесу пошуку стаціонарних точок є найважливішою перевагою і економить обчислювальні ресурси. Багато сучасних методів пошуку сідлових точок засновані на обчисленні і матриці Гессе, зверненні цієї матриці, скалярного добутку вектора градієнта і поточного вектора, знаходженні повного лагранжіан і т.п. Однак всі ці операції є обчислювально «дорогими» і мало б сенс обходити такі складні розрахунки. Ідея модифікації звичайних градієнтних методів, використана в статті, полягає в застосуванні схем пошуку нерухомих точок нелінійних дискретних динамічних систем для задач градієнтного спуску. Передбачається, що цим нерухомим точкам відповідають нестійкі положення рівноваги, і серед мультиплікаторів кожного положення рівноваги є великі одиниці. Використовуються методи усередненого прогнозуючого контролю. Результати чисельного моделювання та візуалізації наведені у вигляді двох таблиць, де вказані басейни тяжіння кожної стаціонарної точки для кожної схеми, і статистичні дані по швидкостям збіжності.
Ключові слова: чисельні методи пошуку сідлових точок, керовані нелінійні дискретні системи, басейни притягань

A. Smorodin. **Predictive control methods in tasks of searching saddle points.** The article presents new methods for searching critical points of a function of several variables, including saddle points. Such problems are found in various fields of theoretical and practical science, for example, saddle-point construction lens design, machine and deep learning, problems of convex optimization and nonlinear programming (necessary and sufficient conditions for the solution are formulated using saddle points of the Lagrange function and proved in the Kuhn-Tucker theorem. When training neural networks, it is necessary to repeat the training process on large clusters and check the network's trainability at different loss functions and different network depth. Which means that thousands of new calculations are run, where each time the loss function is optimized on large amounts of data. So any acceleration in the process of finding critical points is a major advantage and saves computing resources. Many modern methods of searching saddle points are based on calculating the Hessian matrix, inverting this matrix, the scalar product of the gradient vector and the current vector, finding the full Lagrangian, etc. However, all these operations are computationally "expensive" and it would make sense to bypass such complex calculations. The idea of modifying the standard gradient methods used in the article is to apply fixed-point search schemes for nonlinear discrete dynamical systems for gradient descent problems. It is assumed that these fixed points correspond to unstable equilibrium positions, and there are large units among the multipliers of each equilibrium position. The averaged predictive control methods are used. Results of numerical modeling and visualization are presented in the form of two tables, which indicate basins of attraction for each critical point in each scheme, and statistical data by the convergence rates.
Keywords: numerical methods for finding saddle points, controlled nonlinear discrete systems, basins of attraction

Introduction

Methods for solving minimax problems, which are limited to the search for saddle points, are found in different areas of theoretical and practical science. Among the actual problems we can mention, for example, saddle-point construction lens design [1], machine and deep learning [2], problems of convex optimization and nonlinear programming (necessary and sufficient conditions for the solution are formulated using saddle points of the Lagrange function and proved in the Kuhn-Tucker theorem) and many other problems. When training neural networks, you have to repeat the training process on large clusters and check the network's trainability with different loss functions and different network depth, which means that you have to run thousands of new calculations, where each time the loss

DOI: 10.15276/opu.3.62.2020.10

function is optimized on large amounts of data. Thus, any acceleration in the process of finding extremum points is a major advantage and saves computing resources.

Analysis of recent publications and problem statement

Many modern methods in searching saddle points are based on calculating the Hessian matrix, inverting this matrix, the scalar product of the gradient vector and the current vector, finding the full Lagrangian [3], etc. However, all these operations are computationally “expensive” and it would make sense to bypass such complex calculations.

This paper proposes a new approach to solving the considered extreme problems. The idea is to use methods for finding fixed points of nonlinear discrete dynamical systems for gradient descent problems. It is assumed that these fixed points correspond to unstable equilibrium positions. This means that among the multipliers of each equilibrium position there are those of which modulus is greater than one. Moreover, in the case of the gradient descent method, the Jacobian matrix is symmetric, which means that all its multipliers are real. If the fixed point is saddle, then there are positive large units among the multipliers. This case is the most complex in the theory of fixed points’ stabilization. In general, the problem of finding fixed points is limited to the problem of local stabilization of equilibrium positions. To solve these problems, there were proposed various control schemes [4–7], which can be divided into two large groups: methods using the Jacobian matrix or without it. Naturally, it is assumed that the Jacobian matrix at the fixed points themselves is not known, otherwise it would be possible to use the entire powerful apparatus of linear control theory applied to systems linearized in the vicinity of the cycle. The Jacobian matrix is a necessary attribute of Newton-Raphson type methods [8, 9]. Among the methods where the Jacobian matrix is not used, the predictive control method has shown sufficient efficiency [10, 11], which in various special cases allows some modifications.

The purpose of the presented work is to apply different modifications of the method of predictive control stabilization and search for fixed points to searching for saddle points.

The article is organized as follows. Following the Introduction, section “Mathematical foundations of algorithms” provides a brief overview of the results related to the predictive control method for searching the fixed points. In the next section, “Modifying schemes of gradient descent”, the right part of the nonlinear discrete system is replaced by a function of the usual gradient method, and after that the predictive control applies, resulting in the different new scheme of gradient descent. In the section “Numerical modeling results” modification of gradient descent methods are applied to searching saddle points of the model function. The algorithm of calculation, visualization of results are given, as well as their comparison based on the convergence rate and size of basins of attraction the section “Conclusions” examines the issue of the methods effectiveness in general, and the issue of improving this efficiency.

Mathematical foundations of algorithms

A nonlinear discrete system is given:

$$x_{n+1} = f(x_n),\ x_n \in R^m,\ n = 1, 2, \dots, \tag{1}$$

where $f(x)$ – differentiable vector function of the corresponding dimension. It is assumed that this system has one or more unstable fixed points, i.e. $\eta = f(\eta)$. The multipliers of the considered unstable equilibrium positions are defined as the eigenvalues of the Jacobian matrices $f'(\eta)$ of dimensions $m \times m$ at fixed points. Since the fixed points are not known, the Jacobian matrix spectrum is also unknown. For the equilibrium position $x_n = \eta$ of the system (1) the spectrum of the Jacobian matrix is denoted as $\{\mu_1, \dots, \mu_m\}$. We will assume that some set estimate M of the multipliers localization $\{\mu_1, \dots, \mu_m\}$ is known.

Let us take a closer look at the control system:

$$x_{n+1} = F(x_n), \tag{2}$$

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

where $F(x)=\sum_{j=1}^{N}\vartheta_j f^{(j)}(x)$, $f^{(1)}(x)=f(x)$, $f^{(k)}(x)=f(f^{(k-1)}(x))$, $k=2,\dots,N$. Numbers $\vartheta_1,\dots,\vartheta_N$ are real. It is easy to check that with $\sum_{j=1}^{N}\vartheta_j=1$ within the system (2) there is also an equilibrium position η. The task is to select a parameter N and coefficients $\vartheta_1,\dots,\vartheta_N$ so that the equilibrium position η of the system (2) is local to asymptotically stables. Naturally, when constructing these coefficients, it is necessary to use information about the set of localization multipliers M.

Lemma 1 [11]. The Jacobian matrix of the equilibrium position η of system (2) can be conceived of as:

$$\sum_{j=1}^{N}\vartheta_j\ J^j\ , \tag{3}$$

where J – Jacobian matrix of the equilibrium position η of system (1).

Instead of system (2), another control system can be considered:

$$x_{n+1}=f\left(\vartheta_1 x_n+\sum_{j=2}^{N}\vartheta_j f^{(j-1)}(x_n)\right). \tag{4}$$

In case if $\sum_{j=1}^{N}\vartheta_j=1$ is in the system (4), the equilibrium position is maintained. In addition, the Jacobian matrix of this equilibrium position is expressed in terms of the Jacobian matrix of the same equilibrium position of system (1) by the formula (3).

The advantage of the control system (4) over the system (2) is a less number of calculations of the function values $f(x)$. All results for system (2) are transferred without change to system (4).

Theorem 1 [11]. Let $f\in C^1$ and system (1) has an unstable equilibrium position with multipliers $\{\mu_1,\dots,\mu_m\}$. Then this equilibrium position will be locally asymptotically stable of the equilibrium position (2) if:

$$r(\mu_j)\in D,\ \ j=1,\dots,m,$$

where $D=\{z\in C:|z|<1\}$ – open central unit circle, $r(\mu)=\sum_{j=1}^{N}\vartheta_j\ \mu^j$.

Note that the condition $r(1)=1$ must be met.

Different estimates for multipliers allow us designing control systems that stabilize equilibrium positions.

Case A: $M=\{\mu_1,\dots,\mu_m\}$.

If the multipliers are known exactly, then we may choose $N=m+1$ and coefficients $\vartheta_1,\dots,\vartheta_{m+1}$ from the condition $r(\mu)=\sum_{j=1}^{m+1}\vartheta_j\ \mu^j=\frac{\mu}{\prod_{k=1}^{m}(1-\mu_k)}\prod_{k=1}^{m}(\mu-\mu_k)$.

If the starting point belongs to the basin of attraction of the equilibrium position, then the convergence to the cycle is superlinear. This follows from the fact that all multipliers of the equilibrium position of system (2) are equal to zero.

Case B: $M=(-\mu^*,\mu^*)$, $\mu^*>1$.

It follows from Theorem 1 that in order to stabilize the equilibrium position, it is necessary to construct a polynomial $r(\mu)$, so that $r(1)=1$ and $|r(\mu)|\le 1$ for all $|\mu|<\mu^*$. The desired polynomial can be constructed using Chebyshev polynomials of the first kind $T_N(x)=\cos(n\arccos x)$.

Theorem 2 [11]. Let $f \in C^1$ and system (1) has an unstable equilibrium position with multipliers $\{\mu_1, \dots, \mu_m\} \subseteq (-\mu^*, \mu^*)$. Let the value N be chosen from the condition $\dfrac{1}{\sin \pi/2N} > \mu^*$, and coefficients $\vartheta_1, \dots, \vartheta_N$ –

$$r(\mu) = \sum_{j=1}^{N} \vartheta_j \ \mu^j = (-1)^{\frac{N-1}{2}} T_N\left(\mu \sin\frac{\pi}{2N}\right),$$

where $T_N(x)$ – Chebyshev polynomial of the first kind of an odd order N. Then this equilibrium position will be a locally asymptotically stable equilibrium position of system (2) (excluding a finite number of cases, when $\mu_j = (\cos \pi k/N)/(\sin \pi/2N)$, $k = 1, \dots, N-1$).

Case C: $M = (0, \mu^*)$, $\mu^* > 1$.

Theorem 3. Let $f \in C^1$ and system (1) has an unstable equilibrium position with multipliers $\{\mu_1, \dots, \mu_m\} \subseteq (0, \mu^*)$. Let the value N be chosen from the condition $\dfrac{1}{1 + 2\cos \pi/2N} \cot^2 \dfrac{\pi}{4N} > \mu^*$, and coefficients $\vartheta_1, \dots, \vartheta_N$ –

$$r(\mu) = \sum_{j=1}^{N} \vartheta_j \ \mu^j = -T_N\left(\mu\left(\cos\frac{\pi}{N} - \cos\frac{\pi}{2N}\right) + \cos\frac{\pi}{2N}\right),$$

where $T_N(x)$ – Chebyshev polynomial of the first kind of an order N. Then this equilibrium position will be a locally asymptotically stable equilibrium position of system (2) (excluding a finite number of cases).

Case D: $M = (-\mu^*, 1)$, $\mu^* > 1$.

Theorem 4. Let $f \in C^1$ and system (1) has an unstable equilibrium position with multipliers $\{\mu_1, \dots, \mu_m\} \subseteq (-\mu^*, 1)$. Let the value N be chosen from the condition $\cot^2 \dfrac{\pi}{4N} > \mu^*$, and coefficients $\vartheta_1, \dots, \vartheta_N$ –

$$r(\mu) = \sum_{j=1}^{N} \vartheta_j \ \mu^j = T_N\left(\mu\left(1 - \cos\frac{\pi}{2N}\right) + \cos\frac{\pi}{2N}\right),$$

where $T_N(x)$ – Chebyshev polynomial of the first kind of an order N. Then this equilibrium position will be a locally asymptotically stable equilibrium position of system (2) (excluding a finite number of cases).

Lemma 2 [12 – 14]. Let $f \in C^1$ and system (1) has an unstable equilibrium position with multipliers $\{\mu_1, \dots, \mu_m\} \subseteq (-\mu^*, 1)$. Let the value ν meets condition $\dfrac{-1 + \mu^*}{1 + \mu^*} < \nu < 1$. Then this equilibrium position will be a locally asymptotically stable equilibrium position of system:

$$x_{n+1} = (1 - \nu) f(x_n) + \nu x_n.$$

Lemma 2 can be used to solve the stabilization problem in case 2.2. Functions $(-1)^{\frac{N-1}{2}} T_N\left(\mu \sin\dfrac{\pi}{2N}\right)$ if N is odd depict an interval $(-\infty, \mu^*)$ in $(-c^*, 1]$. Function $-T_N\left(\mu\left(\cos\dfrac{\pi}{N} - \cos\dfrac{\pi}{2N}\right) + \cos\dfrac{\pi}{2N}\right)$ if N is odd depicts an interval $(-\infty, \mu^*)$ in $(-c^*, 1]$, and if N is odd $(-\infty, \infty)$ в $(-c^*, 1]$. This means that multipliers of the equilibrium position of system (2) when

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

choosing the coefficients, as indicated in Theorem 2 or 3, fall in the interval $(-c^*, 1]$. Thus, multipliers of the equilibrium position in system:

$$x_{n+1} = (1-\nu)F(x_n) + \nu\, x_n ,$$

where $F(x) = \sum_{j=1}^{N} \vartheta_j f^{(j)}(x)$ or $F(x) = f\left(\vartheta_1 x + \sum_{j=2}^{N} \vartheta_j f^{(j-1)}(x)\right)$, fall within the interval $(-1, 1]$. That is, the equilibrium position becomes locally asymptotically stable (with the possible exception of a finite number of cases).

In a special case, when $N = 2$ $\;-T_N\left(\mu\left(\cos\frac{\pi}{N} - \cos\frac{\pi}{2N}\right) + \cos\frac{\pi}{2N}\right) = 2x - x^2$, if $N = 3$ $(-1)^{\frac{N-1}{2}} T_N\left(\mu \sin\frac{\pi}{2N}\right) = \frac{3}{2}x - \frac{1}{2}x^3$.

Case E: (general case).

Using the ideas of Theorem 1, we can propose the following scheme for stabilizing the equilibrium position. In this case, the coefficients ϑ_j will not necessarily be constants:

a) we determine the matrix $f'(x)$;

b) we determine the characteristic polynomial of this matrix $\sum_{j=1}^{m+1} \vartheta_j(x)\, \mu^{j-1}$;

c) we normalize the characteristic polynomial $\frac{1}{\sum_{j=1}^{m+1} \vartheta_j(x)} \sum_{j=1}^{m+1} \vartheta_j(x)\, \mu^{j-1}$;

d) we build the control system

$$x_{n+1} = F(x_n),$$

where $F(x) = \frac{1}{\sum_{j=1}^{m+1} \vartheta_j(x)} \sum_{j=1}^{m+1} \vartheta_j(x) f^{(j)}(x)$ or $F(x) = f\left(\frac{1}{\sum_{j=1}^{m+1} \vartheta_j(x)}\left(\vartheta_1 x + \sum_{j=2}^{m+1} \vartheta_j(x) f^{(j-1)}(x)\right)\right)$.

Modification schemes of the gradient descent

Let us consider the problem of searching critical points of a function $z = \Phi(\xi_1, \ldots, \xi_m)$, for which it is necessary to solve a system of equations $grad\,\Phi = 0$, where $grad\,\Phi = \left(\frac{\partial\Phi}{\partial\xi_1}, \ldots, \frac{\partial\Phi}{\partial\xi_m}\right)^T$ (symbol "*T*" indicates the transposition operation). Hereafter, we will denote $grad\,\Phi = G(x)$, where $x = (\xi_1, \ldots, \xi_m)^T$. Hessian matrix $H = \left\{\frac{\partial^2\Phi}{\partial\xi_i \partial\xi_j}\right\}_{i,j=1}^{m,m}$ is symmetrical, hence, its eigenvalues $\{\tau_1, \ldots, \tau_m\}$ are real. If these eigenvalues calculated at a critical point of different signs, then the critical point in question is a saddle point, otherwise an extreme one. We consider the general case of eigenvalue signs. A critical point is a fixed point of displaying $x - \gamma G(x)$, where γ – some nonzero number. To search a critical point, let us consider the following iterative scheme:

$$x_{n+1} = x_n - \gamma G(x_n). \qquad (5)$$

That is, we can now consider the problem of searching a fixed point of system (1) in which $f(x) = x - \gamma G(x)$, and apply all schemes suggested in the previous section. The fixed point multipliers are related to the eigenvalues of the Hessian matrix by formulas:

$$\mu_j = 1 - \gamma \tau_j, \; j = 1, \ldots, m.$$

In case if among τ_j there are numbers of different signs, then among the multipliers of the cycle there will necessarily be large units.

Let us assume $\tau_1 \le \ldots \le \tau_m$, at that $\tau_1 < 0$, $\tau_m > 0$. Then if γ is positive, we get $\mu_1 = 1 - \gamma\tau_1 > 1$, $\mu_m = 1 - \gamma\tau_m < 1$. Let's additionally $\mu_m < 0$, and $\mu^* = \max\{\mu_1, |\mu_m|\}$. Then, $\mu_j \in [-\mu^*, \mu^*]$, and we may apply Theorem 2 to build a control scheme. When we choose $N = 3$, then the control scheme (2) will be as follows $x_{n+1} = \frac{3}{2} f(x_n) - \frac{1}{2} f(f(f(x_n)))$, and (4) $x_{n+1} = f\left(\frac{3}{2} x_n - \frac{1}{2} f(f(x_n))\right)$. Let us write down these schemes for the system (5):

Scheme 1.1

$$\begin{cases} u_n = x_n - \gamma G(x_n); \\ v_n = u_n - \gamma G(u_n); \\ w_n = v_n - \gamma G(v_n); \\ x_{n+1} = \frac{3}{2} u_n - \frac{1}{2} w_n. \end{cases}$$

Scheme 1.2

$$\begin{cases} u_n = x_n - \gamma G(x_n); \\ v_n = u_n - \gamma G(u_n); \\ w_n = \frac{3}{2} x_n - \frac{1}{2} v_n; \\ x_{n+1} = w_n - \gamma G(w_n). \end{cases}$$

Theorem 5. Let $G \in C^1$ and parameter γ is chosen from condition $\max\{1 - \gamma\tau_1, |1 - \gamma\tau_m|\} < 2$. Then, the fixed point $x_n = \eta$ of difference schemes 1.1 and 1.2 will be of locally stable equilibrium position.

Now we will apply *Theorem* 3. Let us assume that $\tau_1 \le \ldots \le \tau_m$, where $\tau_1 < 0$, $\tau_m > 0$ and $\mu_m = 1 - \gamma\tau_m \ge 0$. Then, $\mu_j \in [0, 1 - \gamma\tau_1]$, $j = 1, \ldots, m$. We choose $N = 2$, then the control scheme (2) will be as follows $x_{n+1} = 2f(x_n) - f(f(x_n))$, and (4) $x_{n+1} = f(2x_n - f(x_n))$. Let us write down these schemes for the system (5):

Scheme 2.1

$$\begin{cases} u_n = x_n - \gamma G(x_n); \\ v_n = u_n - \gamma G(u_n); \\ x_{n+1} = 2u_n - v_n. \end{cases}$$

Scheme 2.2

$$\begin{cases} u_n = x_n - \gamma G(x_n); \\ v_n = 2x_n - u_n; \\ x_{n+1} = v_n - \gamma G(v_n). \end{cases}$$

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

Theorem 6. Let $G \in C^1$ and parameter γ satisfy the condition $\gamma < \frac{\sqrt{2}-1}{|\tau_1|}$. Then, the fixed point $x_n = \eta$ of difference schemes 2.1 and 2.2 will be of locally stable equilibrium position.

We additionally apply Lemma 2. Now, we get schemes:

Scheme 3.1

$$\begin{cases} u_n = x_n - \gamma G(x_n); \\ v_n = u_n - \gamma G(u_n); \\ x_{n+1} = (1-\nu)(2u_n - v_n) + \nu x_n. \end{cases}$$

Scheme 3.2

$$\begin{cases} u_n = x_n - \gamma G(x_n); \\ v_n = 2x_n - u_n; \\ x_{n+1} = (1-\nu)(v_n - \gamma G(v_n)) + \nu x_n. \end{cases}$$

Theorem 7. Let $G \in C^1$ and parameters γ and ν are chosen from conditions $\gamma < \frac{1}{|\tau_1|}$, $\frac{2\mu - \mu^2 - 1}{2\mu - \mu^2 + 1} < \nu < 1$, where $\mu = 1 - \gamma \tau_m$. Then, the fixed point $x_n = \eta$ of difference systems 3.1 and 3.2 will be of locally stable equilibrium position.

Hereby we note that each given scheme allows us finding all critical points at once, at that parameters must be chosen according to conditions of Theorems. However, the basins of attraction can be very different when moving from one scheme to another. For the effectiveness of the difference schemes application, both knowledge about the basins of attraction of stationary points and information about the convergence rate to a stationary point are important. This is regulated *non-parametrically* by the choice of scheme, *parametrically* – by the choice of an interval and an additional parameter ν.

Results of numerical simulation

Let us consider the test minimax problem [3]:

$$\min_{x \in \Re} \max_{y \in \Re} \left[\Phi(x, y) = 2x^2 + y^2 + 4xy + \frac{4}{3}y^2 - \frac{1}{4}y^4 \right], \tag{6}$$

where $x, y \in \Re$. When solving the equation $grad\, \Phi = 0$, we find all the extremum points:

$$z_0 = (0,0);$$

$$z_1 = (-2+\sqrt{2},\, 2-\sqrt{2});$$

$$z_2 = (-2-\sqrt{2},\, 2+\sqrt{2}).$$

Point z_1 is the local minimum point, but z_0 and z_2 are saddle points.

Let us build the basins of attraction of extreme points of problem (6) for each scheme from the previous section.

Modeling and visualization of numerical results are performed in the Python programming language from the Anaconda distribution package. The graphs are built using the Matplotlib library. All calculations were performed on a computer with Intel(R) Xeon(R) E-2286M CPU @ 2.40GHz with 64GB ECC RAM with the operating system Ubuntu Linux (64 bit) installed.

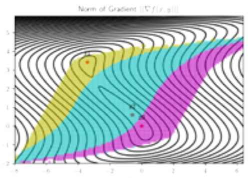
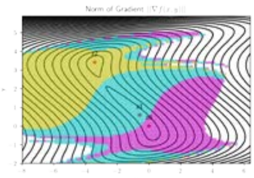
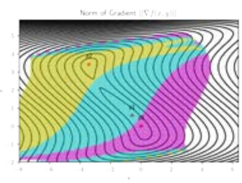
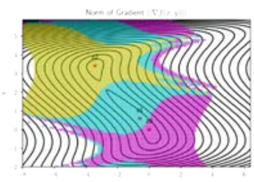
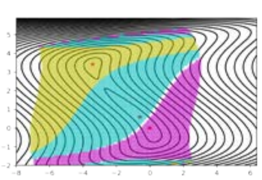
No specialized libraries of increased accuracy were used in the process of calculations, since the stability of methods to standard hardware errors was checked. At the same time, iterative schemes were executed thousands of times based on the finite field arithmetic, which led to rounding the results. There were situations when the result of calculations was the NaN number (Not a number), which meant that results were approaching one of the infinities or went beyond the boundaries of the rectangle in question on the plane (*x*, *y*).

Such cases were treated in a special way and the starting points were not added to the basins of attraction. The grid was built in intervals $x \in [-8.0, 6.0]$ and $y \in [-2.0, 6.0]$ at an interval 0.1. Accuracy has been taken as 10^{-6}. To ensure conditions for comparing different schemes, the same parameters were chosen: $\gamma = 0.1$ and $N = 5000$, in schemes 3.1 and 3.2 additional parameter $\nu = 0.5$.

Basins of attraction for each scheme are shown in Table 1. Purple color indicates the basin of attraction for the point z_0, blue color – z_1, yellow – z_2. The figure shows a significant difference between the basins of attraction when using different schemes. For the experiment conducted, both varieties of scheme 2 were able to converge from a larger number of initial points (64…64.6%). The percentage of convergence was calculated as the percentage of points from the grid, which we used as the starting iteration points, for which there was the convergence to a critical point.

Table 1

Basins of attraction of critical points for difference schemes 1.1, 1.2, 2.1, 2.2, 3.1, 3.2

Scheme name	***Scheme 1.1***	***Scheme 1.2***
basins of attraction visualization		
basins' areas	$z_0 - 1515$ $z_1 - 977$ $z_2 - 2702$	$z_0 - 2057$ $z_1 - 2773$ $z_2 - 2305$
convergence coefficient	~44.77 %	~61.51 %
Scheme name	***Scheme 2.1***	***Scheme 2.2***
basins of attraction visualization		
basins' areas	$z_0 - 2058$ $z_1 - 3022$ $z_2 - 2407$	$z_0 - 2422$ $z_1 - 2192$ $z_2 - 2821$
convergence coefficient	~64.53 %	~64.09 %
Scheme name	***Scheme 3.1***	***Scheme 3.2***
basins of attraction visualization		
basins' areas	$z_0 - 1427$ $z_1 - 2870$ $z_2 - 1321$	$z_0 - 1944$ $z_1 - 1118$ $z_2 - 3878$
convergence coefficient	~48 %	~59.82 %

 ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

To estimate the rate of methods convergence, standard statistical data on this rate were calculated for each of the critical points of the schemes from the different attraction basin points. The data is shown in Table 2, all the values given means the number of iterations spent to approach critical points at a given distance.

Table 2

Statistical characteristics of estimates of the convergence rate to different critical points for different schemes

Scheme	Critical point	Mode	Minimum value	Maximum value	Average value
Scheme 1.1	z_0	585	2	1580	605.96
	z_1	1602	742	2548	1578
	z_2	33	21	88	34.60
Scheme 1.2	z_0	592	2	1509	603
	z_1	1608	807	2818	1573
	z_2	33	21	79	33.42
Scheme 2.1	z_0	920	2	2657	946.02
	z_1	2349	753	3540	2306
	z_2	43	27	123	43.99
Scheme 2.2	z_0	930	2	2699	946.02
	z_1	2345	626	4242	2293
	z_2	43	25	105	42.48
Scheme 3.1	z_0	1845	2	4591	1894.95
	z_1	4690	1795	4997	4573.21
	z_2	90	41	194	93.75
Scheme 3.2	z_1	1852	2	4439	1874.03
	z_2	4696	2297	5000	92.53
	z_3	91	47	214	4544.60

For all the schemes presented in the article, a constant value of the parameter γ was used, which is not optimal from a practical point of view, and many library functions dynamically change this parameter as they approach a critical point. Therefore, these schemes must be modified to take into account an interval change as they approach the critical point.

Conclusions

The methods of predictive control of stabilization and search for fixed points to search for saddle points considered in the paper showed improved results relative to the original method of gradient descent with the new opportunities to find all critical, including saddle points. It should be noted, however, that not for all schemes it was possible to obtain the fastest search for extremals with the maximum number of initial points. This problem will be covered in greater detail in future publications. Additionally, in the following articles it is planned to bring the schemes under consideration to the level of Open Source libraries, such as Scipy (Python). This function has an open extension interface, which it is possible to embed new algorithms in, and this is what we are going to use in the future, and bring all schemes to the level of library ones. This will allow using new methods for solving real-world practical issues, including in the tasks of training neural networks.

Література

1. Bociort F., Maarten van Turnhout, Marinescu O. Practical guide to saddle-point construction in lens design. Proc. SPIE 6667. Current Developments in Lens Design and Optical Engineering VIII, 666708 (18 September 2007). DOI: https://doi.org/10.1117/12.732477.
2. A Fast Saddle-Point Dynamical System Approach to Robust Deep Learning / Esfandiari Y., Balu A., Ebrahimi K, Vaidya U., Elia N., Sarkar S. arXiv:1910.08623.
3. Adolphs L., Hadi D., Lucchi A., Hofmann T. Local Saddle Point Optimization: A Curvature Exploitation Approach. *Proceedings of Machine Learning Research.* 2019. Volume 89. P. 486–495.
4. Ott E., Grebodgi C., Yorke J.A. Controlling chaos. *Phys. Rev. Lett.* 1990. 64. 1196–1199.
5. Chen G., Dong X. From chaos to order: Methodologies, Perspectives and Application. Singapore : World Scientific, 1998, 760 p.
6. Jackson E.A. Perspectives of Nonlinear Dinamics. Vol. I, II. Cambridge Univ. Press, Cambridge, 1980, 1990 Chaos II, ed. Hao Bai-Lin. World Sci., 1990.
7. Yang D., Zhou J. Connections among several chaos feedback control approaches and chaotic vibration control of mechanical systems. *Commun. Nonlinear Sci. Numer. Simulat*. 2014. 19. 3954–3968.
8. Miller J.R., Yorke J.A. Finding all periodic orbits of maps using Newton methods: sizes of basins. *Physica D.* 2000. 135. 195–211.
9. Ypma T.J. Historical Development of the Newton-Raphson Method. *SIAM Rev.* 1995. 37. 531–551.
10. Polyak B.T. Stabilizing chaos with predictive control. *Automation and Remote Control.* 2005. 66 (11). 1791–1804.
11. Dmitrishin D., Iacob E., Stokolos A. Average predictive control for nonlinear discrete dynamical systems. *Advances in Systems Science and Applications.* 2020. 20(1). P. 27–49.
12. Vieira de S.M., Lichtenberg A.J. Controlling chaos using nonlinear feedback with delay. *Phys. Rev.* 1996. E 54. 1200–1207.
13. Morgul O. Further stability results for a generalization of delayed feedback control. *Nonlinear Dynamics*. 2012. 1–8.
14. Dmitrishin D., Skrinnik I., Lesaja G., Stokolos A. (2019). A new method for finding cycles by semilinear control. *Physics Letters A.* 2019. 383, 16. P. 1871–1878. DOI: https://doi.org/10.1016/j.physleta.2019.03.013.

References

1. Bociort, F., Maarten van Turnhout, & Marinescu, O. (2007). *Practical guide to saddle-point construction in lens design, Proc. SPIE 6667.* Current Developments in Lens Design and Optical Engineering VIII. DOI: https://doi.org/10.1117/12.732477.
2. Esfandiari, Y., Balu, A., Ebrahimi, K, Vaidya, U., Elia, N., Sarkar, S. (2020). A Fast Saddle-Point Dynamical System Approach to Robust Deep Learning, arXiv:1910.08623.
3. Adolphs, L., Hadi, D., Lucchi, A., & Hofmann, T. (2019). Local Saddle Point Optimization: A Curvature Exploitation Approach. *Proceedings of Machine Learning Research*, *89*, 486–495.
4. Ott, E., Grebodgi, C., & Yorke, J.A. (1990). Controlling chaos. *Phys. Rev. Lett. 64*, 1196–1199..
5. Chen, G., & Dong, X., (1999). *From chaos to order: Methodologies.* Perspectives and Application. World Scientific, Singapore.
6. Jackson, E.A. (1990). *Perspectives of Nonlinear Dinamics.* Vol. I, II, Cambridge Univ. Press, Cambridge, 1980, 1990 Chaos II, ed. Hao Bai-Lin. World Sci.
7. Yang, D., & Zhou, J. (2014). Connections among several chaos feedback control approaches and chaotic vibration control of mechanical systems. *Commun. Nonlinear Sci. Numer. Simulat, 19*, 3954–3968.
8. Miller, J.R., & Yorke, J.A. (2000). Finding all periodic orbits of maps using Newton methods: sizes of basins. *Physica D, 135*, 195–211.
9. Ypma, T.J. (1995). Historical Development of the Newton-Raphson Method. *SIAM Rev., 37*, 531–551.
10. Polyak, B.T. (2005). Stabilizing chaos with predictive control. *Automation and Remote Control. 66* (11), 1791–1804.

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

11. Dmitrishin, D., Iacob, E., & Stokolos, A. (2020). Average predictive control for nonlinear discrete dynamical systems. *Advances in Systems Science and Applications*, 20(1), 27–49.
12. Vieira de S.M., & Lichtenberg, A.J. (1996). Controlling chaos using nonlinear feedback with delay. *Phys. Rev. E 54*, 1200–1207.
13. Morgul, O. (2012). Further stability results for a generalization of delayed feedback control. *Nonlinear Dynamics*, 1–8.
14. Dmitrishin, D., Skrinnik, I., Lesaja, G., & Stokolos, A. (2019). A new method for finding cycles by semilinear control. *Physics Letters A, 383*, 1871–1878. DOI: https://doi.org/10.1016/j.physleta.2019.03.013.

Смородин Андрій В'ячеславович; Smorodin Andrey, ORCID: https://orcid.org/0000-0002-3370-3197

Received October 14, 2020

Accepted December 11, 2020