A theoretical framework and some promising findings of grey wolf optimizer, part I: analytical model of sampling distribution and stability analysis

Haoxin Wang, Libao Shi*
National Key Laboratory of Power Systems, Shenzhen International Graduate School, Tsinghua University, 518055, P. R. China

Abstract

This paper proposes a theoretical framework of the grey wolf optimizer (GWO) based on several interesting theoretical findings, involving sampling distribution, order-1 and order-2 stability, and global convergence analysis. In the part I of the paper, the characteristics of the sampling distribution of the new solution and the probabilistic stability of the GWO are carefully discussed based on the well-known stagnation assumption for simplification purposes. Firstly, the characteristics of the sampling distribution of the new solution, mainly related to the shape of the joint probability density function (PDF), are discussed under the assumption that the original solution before updating is constant. Then, the assumption that the original solution is constant is eliminated to perform the sampling distribution analysis, based on which several characteristics of the new solution are provided, containing the shape of the joint PDF and central moments of any positive integer order. Finally, the order-1 and order-2 stability of the GWO under stagnation assumption is introduced and proved as the inference of conclusions above, which are all verified by numerical simulations.

Keywords: GWO, sampling distribution, stability analysis, convergence analysis, PDF

1. Introduction

The last few decades have witnessed the flourishing development of a large class of optimization algorithms called meta-heuristic algorithms, which have been regarded as effective solutions to the increasingly complex optimization problems in real life [1-2]. Different from traditional optimization methods like Conjugate Gradient method or Newton method [3], some random searching patterns and mechanisms were introduced in the meta-heuristic algorithms, which substantially reduced the dependence of algorithm performance on the initial solution and increased the probability of finding the global optimal solution of the optimization problem. Moreover, these meta-heuristic algorithms were proposed mainly based on metaphors of some natural or man-made processes, which typified the idea of bionics [4]. Based on the type of natural or man-made process imitated by the meta-heuristic algorithms, they can usually be classified into the following four categories [5]: evolutionary algorithms inspired from Darwin's theory of evolution [6-8], physics-based algorithms inspired from Laws of physics [9-11], human-based algorithms imitating human behavior [12-14], and swarm-based algorithms imitating the behavior of biological populations [15-18]. For more detailed reviews of the meta-heuristic algorithms, please refer to [2, 19-21].
As a typical swarm-based meta-heuristic algorithm, the grey wolf optimizer (GWO) proposed by S. Mirjalili took its metaphor from the hunting process of grey wolf groups [22]. In nature, the grey wolves mostly prefer to live in groups, in which a grey wolf belongs to one of the four categories: α wolf, β wolf, δ wolf, and ω wolf [23]. During predation, the ω wolves are guided by the α wolf, β wolf, and δ wolf to move towards the prey. In the GWO, each grey wolf in the group was regarded as a search agent, and the α wolf, β wolf, and δ wolf were regarded as the best search agent, the second best search agent, and the third best search agent, respectively. Other search agents (representing ω wolves) updated their positions under the guidance of the positions of α wolf, β wolf, and δ wolf (the detailed updating equation of each search agent can be found in section 2). Since its proposal in 2014, owing to its intuitive procedure and outstanding performance, the GWO has attracted widespread attention and has been applied to solve various optimization problems in real life, such as optimal operation of power systems [24-26], training of neural networks [27-28], clustering analysis [29-30], and path planning [31-32], etc. Please refer to [33-35] for more detailed reviews of applications of the GWO.

Despite the research on applications, studies have been done on the performance analysis of the GWO as well. In fact, the updating mechanism and the distribution of particles within the feasible region were already analyzed qualitatively and presented in the form of diagrams in the original literature of the GWO (please see Fig. 3 – Fig. 5 given in [22]). Later on, the convergence behavior of the GWO was verified and studied by numerical simulations based on various test functions [36], and the performance of the GWO was compared with several well-known meta-heuristic algorithms such as the particle swarm optimization (PSO), differential evolution (DE), firefly algorithm (FA), etc. [37-39]. In addition, the mechanisms that how parameters in the updating equation of the GWO controlled the distribution of particles were studied qualitatively, based on which extensive research has been carried out on the improvement of the original GWO from various aspects, such as parameter selection and control [40-41], revision of the updating equation [42-44], and study of population topology [45], etc. It can be seen that till now, most studies on the performance analysis of the GWO were conducted qualitatively via diagrams or numerical simulations, and there were little research on the theoretical foundations of the GWO based on quantitative analysis, such as sampling distribution, probabilistic stability analysis, local/global convergence analysis, etc.

In the part I of the paper, we propose a theoretical framework of the GWO based on several new theoretical findings, involving sampling distribution, order-1 and order-2 stability, and global convergence analysis. In particular, we analyze the characteristics of the sampling distribution of the new solution, and some promising findings of the GWO are carefully discussed, in which the stagnation assumption is introduced and concerned for simplification. Firstly, the original solution before updating is treated as a constant rather than a random variable. Under this assumption, the new solution could be treated as a function of random parameters. Then, the assumption above is eliminated, that is, the original solution is still a random variable. In this case, the solution should be regarded as a stochastic process corresponding to the iteration of the GWO rather than a simple random variable. Finally, as an application of the sampling distribution characteristics of the solution analyzed above, we introduce and prove the order-1 and order-2 stability of the GWO under stagnation assumption. In the part II of the paper, based on the theoretical results concluded in the part I, we will prove the global convergence of the GWO. All of the conclusions above are verified by numerical simulation results.
The part I of the paper is organized as follows. In section 2, the updating mechanism and procedure of the GWO are introduced in brief. In section 3, the stagnation assumption is introduced, and the sampling distribution of the new solution with both stagnation assumption and constant original solution assumption is solved and analyzed. In section 4, the constant original solution assumption is eliminated, the characteristics of the sampling distribution of the new solution is discussed only under the stagnation assumption, and the order-1 and order-2 stability of the GWO under stagnation assumption are proved successively. In section 5, we conclude the theoretical results above.

2. An introduction of GWO

As mentioned above, the GWO algorithm is inspired by the leadership hierarchy and hunting mechanism of grey wolves in nature. In the GWO, the best, the second best, and the third best search agents are called $\alpha$, $\beta$, and $\delta$ wolves, respectively, and other search agents are called $\omega$ wolves. Let $T$ be the total number of iterations of the GWO, $D$ be the dimension of the optimization problem, $N$ be the number of search agents in the group, and the positions of $\alpha$, $\beta$, and $\delta$ wolves at iteration $t$ are marked as

$$ p_1(t) = [p_{11}(t), \ldots, p_{1D}(t)] = [p_{1j}(t)]_{j = 1, \ldots, D}, \quad p_2(t) = [p_{2j}(t)]_{j = 1, \ldots, D}, \quad p_3(t) = [p_{3j}(t)]_{j = 1, \ldots, D}, $$

respectively. The position of the $i$th $\omega$ wolf at iteration $t$ $x_i(t) = [x_{ij}(t)]_{i = 1, \ldots, N; j = 1, \ldots, D}$ is updated by the guidance of the positions of $\alpha$, $\beta$, and $\delta$ wolves $p_k(t) (k = 1, 2, 3)$, which can be formulated as:

$$ x_i(t + 1) = \frac{1}{3} \sum_{k=1}^{3} x'_k(t) \tag{2.1} $$

where $x'_k(t) = [x'_{kj}(t)]_{k = 1, 2, 3; j = 1, \ldots, D}$ is an intermediate vector, and $x'_{kj}(t)$ can be calculated as:

$$ x'_{kj}(t) = p_{kj}(t) + A_{kj}[c_{kj}p_{kj}(t) - x_{ij}(t)] \tag{2.2} $$

where $A_{kj} \sim U[-a(t), a(t)]$ ($U[p, q]$ denotes the uniform distribution with lower bound $p$ and upper bound $q$, similarly hereinafter), $c_{kj} \sim U[0, 2]$, and $a(t)$ is linearly decreased from 2 to 0 during iteration:

$$ a(t) = 2 \left(1 - \frac{t}{T}\right) \tag{2.3} $$

Obviously, $x'_1(t), x'_2(t),$ and $x'_3(t)$ represent the guidance of the positions of $\alpha$, $\beta$, and $\delta$ wolves to that of the $i$th $\omega$ wolf, respectively. According to Eq. (2.1), the new position of the $i$th search agent $x_i(t + 1)$ is an average of the three intermediate vectors $x'_k(t) (k = 1, 2, 3)$, which provides an intuitive reflection of the hunting process of $\omega$ wolves in real life. In addition, when the fitness of the new search agent $x_i(t + 1)$ is better than that of the best three search agents at iteration $t$ $p_k(t)$, the new position of the best three search agents $p_k(t + 1)$ will be replaced by $x_i(t + 1)$, which can be formulated as:

$$ p_1(t + 1) = \begin{cases} x_i(t + 1), f(x_i(t + 1)) < f(p_1(t)) \\ p_1(t), \text{otherwise} \end{cases} \tag{2.4} $$

$$ p_2(t + 1) = \begin{cases} x_i(t + 1), f(p_2(t)) \leq f(x_i(t + 1)) < f(p_2(t)) \\ p_2(t), \text{otherwise} \end{cases} \tag{2.5} $$

$$ p_3(t + 1) = \begin{cases} x_i(t + 1), f(p_3(t)) \leq f(x_i(t + 1)) < f(p_3(t)) \\ p_3(t), \text{otherwise} \end{cases} \tag{2.6} $$

where the optimization problem is assumed to be $\min f(x)$. In summary, the pseudo-code of the GWO is given in Fig. 1.
3. Analytical model of sampling distribution of $x_i(t+1)$ with constant $x_i(t)$ assumption

In this section, the stagnation assumption is introduced and explained under the assumption that $x_i(t)$ is treated as a constant, and the sampling distribution characteristics of the positions of the intermediate vector $x'_k(t)$ as well as the new solution $x_i(t+1)$ are derived and analyzed successively.

3.1 Preliminaries and assumptions

According to Eqs. (2.1) - (2.3), the position of the $i$th search agent at iteration $t + 1$ (marked as $x_i(t + 1)$) is a function of the position of the $i$th search agent at iteration $t$ $x_i(t)$, the position of the best three search agents at iteration $t$ $p_k(t)$ ($k = 1, 2, 3$), and the random parameters $A_{kj}, C_{kj}$. Obviously, the position vectors $x_i(t)$ and $p_k(t)$ are constantly changing during iterations, thus the $x_i(t)$ should be seen as a complicated stochastic process corresponding to the iteration $t$, which brings a certain degree of difficulty to the analysis and solution of the sampling distribution characteristics of $x_i(t+1)$. Accordingly, in view of the convenience of problem analysis, it is necessary to put forward some appropriate assumptions.

In the part I of the paper, the stagnation assumption is introduced and applied to the theoretical analysis of the GWO, i.e., the positions of the best three search agents $p_k(t)$ keep unchanged during iterations (as a result, in this section and the next section, the $p_k(t)$ will be abbreviated as $p_k$). As a common assumption in the field of the analysis of meta-heuristic algorithms, the stagnation assumption has been widely applied to simplify the sampling distribution and stability analysis of the particle swarm optimizer (PSO), in which the global best (gbest) of the population and the personal best (pbest) of each search agent are set to be constant [46-47]. Obviously, in the GWO, the stagnation assumption is usually approximately satisfied at the later stage. More
specifically, in the earlier stage of the GWO (please see Fig. 2(a)), since the search agents are distributed in the whole feasible region of the optimization problem, the searching space of each agent is relatively large, and the population holds a wide searching range (represented by the big cube $D_i$ as shown in Fig. 2(a)), thus the ability to search for the neighborhood of the global optimum (represented by the small cube $D_g$ and point $x_g$ as shown in Fig. 2(a)) is guaranteed. However, in the later stage of the GWO, the stagnation assumption is approximately satisfied, i.e. the relationship $p_1 \approx p_2 \approx p_3$ holds, and the positions of all other search agents are close to that of $p_k$. In the optimistic scenarios, the $p_k$ is within a small neighborhood of the global optimum, which means the GWO succeeds in finding the global optimum; however, in the pessimistic scenarios (please see Fig. 2(b)), the $p_k$ is quite far from the global optimum, thus the ability of the population to search for the global optimum is no longer guaranteed.

With the consideration of stagnation assumption, the positions of the best three search agents can be seen independent of iteration $t$, however, the position of search agent $i$ $x_i(t)$ is still a stochastic process corresponding to $t$. In this section, we further assume that the $x_i(t)$ is a constant vector (thus abbreviated as $x_i$ in this section), in which case the $x_i(t+1)$ is no longer a stochastic process, but merely a random variable calculated as (according to Eqs. (2.1) and (2.2)):

$$x_{ij}(t+1) = \frac{1}{3} \sum p_j + \frac{1}{3} \sum A_{kj} C_{kj} p_k - x_{ij}$$

(3.1)

where $\sum p_j = \sum_{k=1}^3 p_{kj}$, similarly hereinafter.

According to Eq. (3.1), the $x_{ij}(t+1)$ becomes a function of six random variables $A_{kj}, C_{kj}(k = 1, 2, 3)$, thus the difficulty of analyzing the sampling distribution characteristics of $x_i(t+1)$ is substantially reduced.

**Remark 1**

In order to better understand the work to be done in this section, a simplified solution updating equation applied to evolution strategy (ES) [48] and evolutionary programming (EP) [49] algorithm as given in (3.2) is utilized to further explain our purpose:

$$x_{ij}(t+1) = x_{ij}(t) + N(0, \sigma^2)$$

(3.2)

where $N(0, \sigma^2)$ denotes a random variable following a normal distribution with mean 0 and variance $\sigma^2$. Although the distribution of the stochastic process $x_{ij}(t)$ is hard to derive, we can
easily conclude that the new solution \( x_{ij}(t + 1) \) follows a normal distribution with mean \( x_{ij}(t) \) and variance \( \sigma^2 \) under the assumption that \( x_{ij}(t) \) is treated as a constant, which points out an obvious direction for researchers to control the mutation scale of the ES/EP by adjusting the parameter \( \sigma \) [50]. Furthermore, extensive research has been carried out to improve the performances of the ES and EP by changing the \( N(0, \sigma^2) \) given in Eq. (3.2) into random variables following other kinds of distributions (such as Cauchy distribution [51], Levy distribution [52], etc. [53]). All of the research above show the significance of studying the sampling distribution characteristics of the new solution \( x_{ij}(t + 1) \) with assumption that the original solution \( x_{ij}(t) \) is treated as a constant. As shown in Eqs. (2.1) and (2.2), the solution updating equation of the GWO is designed by straightly mimicking the predation behavior of grey wolves rather than providing the distribution types (or other detailed characteristics) of the updated solutions, therefore in this section we aim at analyzing and modeling the sampling distribution characteristics of the updated solution under the stagnation and constant \( x_{ij}(t) \) assumptions.

According to Eq. (2.1), the \( x_{ij}(t + 1) \) is a linear combination of three items with the same form \( x_k'(t) \) (abbreviated as \( x_k' \) in this section), which means the joint probability density function (PDF) of \( x_{ij}(t + 1) \) can be derived by calculating the convolution among those of \( x_k'(k = 1, 2, 3) \). Thus, we will firstly analyze and derive the sampling distribution characteristics of \( x_k' \) in section 3.2, and then discuss the sampling distribution characteristics of \( x_{ij}(t + 1) \) in section 3.3.

### 3.2 Sampling distribution analysis of \( x_k' \)

In this section, under the stagnation and constant \( x_{ij}(t) \) assumptions, the sampling distribution of \( x_k' \) is firstly derived and discussed. According to Eq. (3.1), each component of the new solution \( x_{ij}(t + 1)(j = 1, ..., D) \) is calculated independently. In addition, all of the 2D random parameters \( A_{kj}, C_{kj}(j = 1, ..., D) \) are independent of each other, thus all of the \( D \) random variables \( x_{ij}(t + 1)(j = 1, ..., D) \) are independent of each other. Consequently, let the joint PDF of \( x_k' \) be \( c_{kj}(\mathbf{u}) = \prod_{j=1}^{D} g_{kj}(u_j) \) and the PDF of \( x_k' \) be \( g_{kj}(u_j) \), we have (all necessary proofs of propositions and corollaries can be found in Appendix):

**Proposition 1.1**

\[
    c_{kj}(\mathbf{u}) = \prod_{j=1}^{D} g_{kj}(u_j) \tag{3.3}
\]

where the expression of \( g_{kj}(u_j) \) can be derived as follows:

**Proposition 1.2**

Let

\[
    m_{kj} = a(t)(-|p_{kj}| + |x_{ij} - p_{kj}|), n_{kj} = a(t)(|p_{kj}| + |x_{ij} - p_{kj}|) \tag{3.4}
\]

When \( m_{kj} \leq 0 \),

\[
    g_{kj}(u_j) = \begin{cases} 
        \frac{1}{n_{kj}-m_{kj}} \ln \frac{\sqrt{m_{kj}/|u_j-p_{kj}|}}{|u_j-p_{kj}|}, & |u_j-p_{kj}| < -m_{kj} \\
        \frac{1}{2(n_{kj}-m_{kj})} \ln \frac{n_{kj}}{|u_j-p_{kj}|}, & -m_{kj} \leq |u_j-p_{kj}| < n_{kj} \\
        0, & |u_j-p_{kj}| \geq n_{kj}
    \end{cases} \tag{3.5}
\]

Otherwise,
As the core conclusion of this section, the proposition 1.2 gives an exact analytical expression for the PDF of intermediate variable $x'_{k,j}$. According to the proposition 1.2, obviously we have:

**Corollary 1.1**

The domain of $g_{kj}(u_j)$ is \( \{ u_j \in \mathbb{R} | g_{kj}(u_j) > 0 \} = (p_{kj} - n_{kj}, p_{kj} + n_{kj}) \) (in this corollary, the domain is defined as \( \{ u_j \in \mathbb{R} | g_{kj}(u_j) > 0 \} \), however, in the rest of the paper, for the sake of simplicity, we might say that a certain function $h$ holds some characteristics within $\mathbb{R}$, yet the domain of $h$ as defined in this corollary is only a subset of $\mathbb{R}$, and the value of $h$ that exceeds the domain (as defined in this corollary) is equal to 0). The curve of $g_{kj}(u_j)$ is symmetrical about $u_j = p_{kj}$, and $g_{kj}(u_j)$ is non-decreasing within $(p_{kj} - n_{kj}, p_{kj})$.

From corollary 1.1, the distribution of $x'_{k,j}$ is a bounded single-peak distribution symmetrical about $p_{kj}$.

According to the proposition 1.2, the curves of $g_{kj}(u_j)$ under cases that $m_{kj} \leq 0$ and $m_{kj} > 0$ can be drawn as Fig. 3 (a) and Fig. 3 (b), respectively, in which all curve parts correspond to the logarithmic functions. It is worth noting that when $m_{kj} \leq 0$, the curve of $g_{kj}(u_j)$ within $(p_{kj} - n_{kj}, p_{kj})$ is formed by two different logarithmic curves (i.e. $g_{kj}(u_j)$ is not differentiable at $u_j = p_{kj} \pm m_{kj}$). Furthermore, when $m_{kj} \leq 0$, the range of $g_{kj}(u_j)$ is not bounded, and $g_{kj}(p_{kj}) = +\infty$ holds.

![Fig. 3. Curves of $g_{kj}(u_j)$. (a): $m_{kj} \leq 0$, (b): $m_{kj} > 0$.

The proposition 1.2 can be verified by Monte Carlo (MC) simulations. Fig. 4 shows the frequency histogram of $x'_{k,j}$ obtained by MC simulation and its PDF curve given in Eqs. (3.4) – (3.6) with four sets of values of variables $\alpha(\xi), p_{kj}, x_{ij}$, and $m_{kj}$. For each set of variable values, the sample size is set to be $10^6$, and the number of frequency histogram segments is set to be 60. Obviously, as the sample size and the number of frequency histogram segments increase, the frequency histogram of $x'_{k,j}$ obtained by MC simulation tends to coincide with the curve of the PDF of $x'_{k,j}$ (i.e. $g_{kj}(u_j)$). It can be seen from Fig. 4 that whatever the sign of $m_{kj}$ is, the frequency histogram of $x'_{k,j}$ coincides with the analytical curve of $g_{kj}(u_j)$ (as calculated by Eqs. (3.4)-(3.6)) very well.
Fig. 4. Frequency histograms of $x'_{kj}$ and curves of $g_{kj}(u_j)$. (a): $a(t) = 2$, $p_{kj} = 1$, $x_{ij} = 3$, $m_{kj} = 2 > 0$ , (b): $a(t) = 2$, $p_{kj} = 1$, $x_{ij} = 0.5$, $m_{kj} = -1 < 0$ , (c): $a(t) = 0.5$, $p_{kj} = 1$, $x_{ij} = 3$, $m_{kj} = 0.5 > 0$ , (d): $a(t) = 0.5$, $p_{kj} = 1$, $x_{ij} = 0.5$, $m_{kj} = -0.25 < 0$

Based on the analytical model of $g_{kj}(u_j)$ discussed above, in the last part of this section, the distribution characteristics of $x'_{k}$ will be discussed elaborately. Since all of the $D$ random variables $x_{ij}(t+1)(j = 1, ..., D)$ are independent of each other, and according to the corollary 1.1, the domain of $g_{kj}(u_j)$ is $(p_{kj} - n_{kj}, p_{kj} + n_{kj})$, obviously we have:

**Corollary 1.2**

The random vector $x'_{k}$ is distributed in the $D$-dim box:

$$D_k = \prod_{j=1}^{D} (p_{kj} - n_{kj}, p_{kj} + n_{kj})$$

(3.7)

According to the corollary 1.2, the center of $D_k$ is $p_k$, and the length of the $j$th side of $D_k$ is $2n_{kj}$. Obviously, the measurement of $D_k$ is an increasing function of $a(t)$ (please see Eq. (3.4)), which means the searching space of agent $i$ decreases during iterations.

The MC simulation method is applied to illustrate the distribution characteristics of $x'_{k}$. Here, the sample size of $x'_{k}$ is set to be 5000, and $a = 2$, $p_k = [1, 1], x_i = [1.5, 3]$, then the MC simulation results of $x'_{k}$ (marked by blue points) and the theoretical value of $D_k$ (as calculated via Eq. (3.7), marked by blue dashed rectangle) are given in Fig. 5. It can be seen that $D_k$ is approximately an envelope of $x'_{k}$, whose center is $p_k$, and the side of length is directly proportional to $a(t)$ (in Fig. 5 (a), $a(t) = 2$, the size of $D_k$ is $6 \times 12$; in Fig. 5 (b), $a(t) = 0.5$, the size of $D_k$ is $1.5 \times 3$). Furthermore, the probability that $x'_{k}$ appears next to $p_k$ is obviously larger than that next to the boundary of $D_k$, which coincides with the conclusion that $g_{kj}(u_j)$ is a single-peak function (please see the corollary 1.1).
3.3 Analytical modeling of sampling distribution of $x_t(t + 1)$

In section 3.2, we derived the analytical expression and discussed the distribution characteristics that $x'_k$ follows. Based on the results obtained above, in this section, we will study the distribution model that $x_t(t + 1)$ follows under stagnation and constant $x_t(t)$ assumptions. According to Eq. (2.1), the $x_{ij}(t + 1)$ is a linear combination of $x'_{kj}(k = 1, 2, 3)$, thus the PDF of $x_{ij}(t + 1)$ can be derived by calculating the convolution among $g_{kj}(u_j)(k = 1, 2, 3)$ [54]. Let the PDF of $x_{ij}(t + 1)$ be $h_{ij}(u_j)$, obviously we have:

**Proposition 2.1**

$$h_{ij}(u_j) = (g_{1j} * g_{2j} * g_{3j})(3u_j)$$  \hspace{1cm} (3.8)

Although the expression of $h_{ij}(u_j)$ is given in proposition 2.1 in the convolution form, it is almost impossible to substitute Eqs. (3.5) and (3.6) into Eq. (3.8) to derive the exact analytical expression of $h_{ij}(u_j)$. This is because the expression of $g_{kj}(u_j)$ is a piecewise function, hence the exact analytical expression of $h_{ij}(u_j)$ becomes a more complicated piecewise function with more subsections after convolution, and the endpoints of each segment require lengthy and complicated discussions, let alone the difficulty of calculating the exact analytical expression on each segment.

**Remark 2**

In order to better illustrate the complexity of the exact analytical expression of $h_{ij}(u_j)$, we try to derive the expression of $(g_{1j} * g_{2j})(u_j)$ on its leftmost subsection region. We suppose that $m_{1j}, m_{2j} > 0$, and let $g_{10}(u_j) = g_{1j}(u_j + p_{1j}), g_{20}(u_j) = g_{2j}(u_j + p_{2j})$, then both $g_{10}(u_j), g_{20}(u_j)$ are piecewise functions, whose subsection region endpoints are (from left to right) $-n_{1j}, -m_{1j}, m_{1j}, n_{1j}$ and $-n_{2j}, -m_{2j}, m_{2j}, n_{2j}$, respectively, which means the subsection region endpoints of $g_{10} * g_{20}$ are $\pm n_{1j} \pm n_{2j}, \pm m_{1j} \pm m_{2j}$ (there may be duplicates). The number of subsection regions may be as many as 15, let alone the tedious discussion of the relationship among these endpoints.

In addition, we try to derive the expression of the convolution results on its leftmost subsection region. Suppose that $-n_{1j} - m_{2j} < -n_{1j} - m_{2j}$, in which case the leftmost subsection region of $(g_{1j} * g_{2j})(u_j)$ becomes $(-n_{1j} - n_{2j}, -n_{1j} - m_{2j})$, and we have:

$$(g_{1j} * g_{2j})(u_j + p_{1j} + p_{2j})$$
\[
= (g_{10} * g_{20})(u_j)
= \int_{-n_{1j}}^{u_j+n_{2j}} \frac{1}{2(n_{1j}-m_{1j})} \cdot \ln \frac{n_{1j}}{p_{1j}-v} \cdot \frac{1}{2(n_{2j}-m_{2j})} \cdot \ln \frac{n_{2j}}{p_{2j}-v} \, dv
= \frac{1}{4(n_{1j}-m_{1j})(n_{2j}-m_{2j})} \int_{-n_{1j}}^{u_j+n_{2j}} \ln \frac{n_{1j}}{p_{1j}-v} \cdot \ln \frac{n_{2j}}{p_{2j}-v} \, dv
= \frac{1}{4(n_{1j}-m_{1j})(n_{2j}-m_{2j})} \left\{ v - (v - p_{1j})[-1 + \ln(v - p_{1j})] + (v - p_{1j}) \left( 1 + \ln \frac{n_{1j}}{v-p_{1j}} \right) \right\} \ln \frac{n_{2j}}{p_{2j}-v} + \left( \frac{n_{2j}}{p_{2j}-v} \right) \left( p_{2j} + p_{3j} \right) \left[ \ln(v - p_{1j}) \cdot \ln \frac{v-p_{2j}}{p_{1j}-p_{2j}} + L i_{2} \left( \frac{v-p_{2j}}{p_{1j}-p_{2j}} \right) \right] \right\}_{v=-(u_j+n_{2j})}
\]

where \( L i_{n}(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n} \) denotes the poly-logarithm function [55]. Evidently, such a lengthy result has little practical meaning and application value. Taking the tedious classified discussion of subsection region endpoints into account, we could easily conclude that the exact analytical expression of \( h_{ij}(u_j) \) becomes too lengthy to hold any practical value.

Fortunately, based on the results provided in section 3.2, we can further analyze the characteristics of \( h_{ij}(u_j) \) as follows. If setting
\[
\sum p_j = p_{1j} + p_{2j} + p_{3j} \quad \sum m_j = m_{1j} + m_{2j} + m_{3j} \quad \sum n_j = n_{1j} + n_{2j} + n_{3j}
\]
for the sake of simplicity, then we have:

**Corollary 2.1**

\( h_{ij}(u_j) \) is a continuous bounded function defined within \( \{ u_j \in \mathbb{R} | h_{ij}(u_j) > 0 \} = \left( \frac{1}{3} \sum p_j - \frac{1}{3} \sum n_j, \frac{1}{3} \sum p_j + \frac{1}{3} \sum n_j \right) \).

According to the corollary 2.1, the domain and range of \( h_{ij}(u_j) \) can be obtained. Similar with that of \( g_{kij}(u_j) \), the domain of \( h_{ij}(u_j) \) is an open interval with finite length, which means when \( x_{ij}(t) \) is regarded as constant, the searching space of \( x_{ij}(t+1) \) is limited within a range symmetrical about \( \frac{1}{3} \sum p_j \) (different from the mutation operator of the ES/EP as shown in Remark 1, if the mutation operator of ES/EP is applied to update \( x_{ij} \), the searching space of \( x_{ij}(t+1) \) will become the whole feasible region). In addition, it is worth noting that the non-continuous function \( g_{kij}(u_j) \) with unbounded range (when \( m_{kj} \leq 0 \)) is transformed into a continuous function \( h_{ij}(u_j) \) with bounded range after the application of convolution operator, in which process the smoothing effect of convolution operator [56-57] is demonstrated, which will be elaborated in corollary 2.3.

**Corollary 2.2**

The curve of \( h_{ij}(u_j) \) is symmetrical about \( u_j = \frac{1}{3} \sum p_j \), \( h_{ij} \left( \frac{1}{3} \sum p_j - \frac{1}{3} \sum n_j \right) = 0 \), and \( h_{ij}(u_j) \) is non-decreasing within \( \left( \frac{1}{3} \sum p_j - \frac{1}{3} \sum n_j, \frac{1}{3} \sum p_j \right) \).
According to the corollary 2.2, after convolution, the symmetry and unimodal characteristics of $g_k(u_j)$ are both retained, i.e. the $h_{ij}(u_j)$ is a single-peak function symmetrical about $u_j = \frac{1}{3} \sum p_j = \frac{1}{3} \sum_{k=1}^{3} p_{kj}$, which is exactly the average of those of $g_{k_j}(u_j)(k = 1,2,3)$ (i.e. $p_{kj}$).

**Corollary 2.3**

When there exists at least one positive number among $m_{kj}(k = 1,2,3)$, the $h_{ij}(u_j)$ is derivable within $\mathbb{R}$, and $g'_{ij}\left(\frac{1}{3} \sum p_j\right) = 0$ holds.

The corollary 2.3 provides a further indication of the smoothing effect of convolution operator [56-57]. According to the proposition 1.2 (as shown in Fig. 3), the $g_{k_j}(u_j)$ is non-differentiable at $u_j = p_{kj} \pm m_{kj}$, and when $m_{kj} \leq 0$, $g_{k_j}(p_{kj}) = +\infty$, the $g_{k_j}(u_j)$ is non-continuous at $u_j = p_{kj}$. However, when there exists at least one positive number among $m_{kj}(k = 1,2,3)$, the smoothing effect of convolution operator makes $h_{ij}(u_j)$ a differentiable function within $\mathbb{R}$. However, the smoothing effect is limited within conditions that $\exists k, m_{kj} > 0$. When $\forall k = 1,2,3, m_{kj} \leq 0$, $g_{k_j}(u_j)(k = 1,2,3)$ are all non-continuous functions, then the differentiability of $h_{ij}(u_j)$ at $u_j = \frac{1}{3} \sum p_j$ is no more guaranteed.

The conclusions above can be verified by MC simulations. Here, the sample size is set to be $10^6$, and the number of frequency histogram segments is set to be 60. Since the sign of $m_{kj}$ has a significant impact on the shape of the curves of $g_{k_j}(u_j)$ and $h_{ij}(u_j)$, the simulations are carried out under the following four cases as shown in Eq. (3.10). In addition, in each case two groups of values of $x_{ij}$ and $p_{kj}$ are randomly chosen, thus 8 groups of the frequency histograms of $x_{ij}(t + 1)$ are generated and shown in Fig. 6. In order to better exhibit the characteristics of $h_{ij}(u_j)$ generated by MC simulations, the upper-side midpoints of all rectangles in each frequency histogram are connected together as the curve of $h_{ij}(u_j)$. For reader’s reference, all parameters corresponding to each frequency histogram are given in Table 1.

$$\sum_{k=1}^{3} \epsilon(m_{kj}) = 0,1,2,3, \text{where } \epsilon(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$  

(3.10)
Fig. 6. Frequency histograms, please see Table I for parameters corresponding to each sub-graph.

Table I Parameters corresponding to each sub-graph as shown in Fig. 6.

| Subfigure No. | \( p_{1j} \) | \( p_{2j} \) | \( p_{3j} \) | \( \sum_{k=1}^{3} p_{kj} \) | \( x_{ij} \) | \( \sum_{k=1}^{3} e(m_{kj}) \) |
|---------------|--------------|--------------|--------------|----------------|-------------|----------------|
| a             | 0.43         | 1.74         | 2.70         | 1.62           | 0.83        | 0              |
| b             | -0.18        | -0.97        | -1.31        | -0.82          | -0.22       | 0              |
| c             | 0.57         | 0.86         | -1.56        | -0.04          | 0.72        | 1              |
| d             | 0.65         | 1.13         | -2.89        | -0.37          | 0.90        | 1              |
| e             | 0.09         | -1.07        | 2.88         | 0.63           | 2.08        | 2              |
| f             | 0.56         | -1.85        | -0.34        | -0.54          | -3.14       | 2              |
| g             | -0.30        | 0.96         | 2.24         | 0.97           | -2.87       | 3              |
| h             | 0.22         | 0.55         | 0.18         | 0.32           | 3.11        | 3              |

It can be seen from Fig. 6 that whatever the values of \( x_{ij} \) and \( p_{kj} \) are, the \( h_{ij}(u_j) \) are
always single-peak functions symmetrical about \( u_j = \frac{1}{3} \sum p_j \). In addition, when there exists at least one positive number among \( m_{kj}(k = 1, 2, 3) \) (corresponding to Fig. 6 (c) – (h)), the curve of \( h_{ij}(u_j) \) is smooth, and the \( h_{ij}(u_j) \) is derivable within \( \mathbb{R} \); when \( \forall k = 1, 2, 3, m_{kj} \leq 0 \) holds (corresponding to Fig. 6 (a) – (b)), the curve of \( h_{ij}(u_j) \) is not guaranteed to be smooth at \( u = \frac{1}{3} \sum p_j \), which means the \( h_{ij}(u_j) \) might not be derivable at \( u_j = \frac{1}{3} \sum p_j \). Thus, corollaries 2.1 – 2.3 are verified.

In the last part of this section, the distribution characteristics of \( x_i(t + 1) \) is discussed. Since all of the \( D \) random variables \( x_{ij}(t + 1)(j = 1, ..., D) \) are independent of each other, let the joint PDF of \( x_i(t + 1) \) be \( c_i(u) \), then we have:

**Proposition 2.2**

\[
c_i(u) = \prod_{j=1}^{D} h_{ij}(u_j) \tag{3.11}
\]

Based on the analytical expression and characteristics of \( g_{ij}(u_j) \) discussed above, we have:

**Corollary 2.4**

The random vector \( x_i(t + 1) \) is distributed in the \( D \)-dim box:

\[
D_i = \prod_{j=1}^{D} \left( \frac{1}{3} \sum p_j - \frac{1}{3} \sum n_j, \frac{1}{3} \sum p_j + \frac{1}{3} \sum n_j \right) \tag{3.12}
\]

According to the corollary 2.4, the center of \( D_i \) is \( \frac{1}{3} \sum p \), and the length of the \( j \)-th side of \( D_i \) is \( \frac{2}{3} \sum n_j \). Similar to the characteristics of \( D_k \) (as given in the corollary 1.2), the measurement of \( D_i \) is an increasing function of \( a(t) \) (please see Eq. (3.9)), which means the searching space of agent \( i \) decreases during iterations. Here, the MC simulation is applied to illustrate the distribution of \( x_i(t + 1) \). The sample size of \( x_i(t + 1) \) is set to be 5000, and \( D = 2, p_1 = [0, 0], p_2 = [0, 1], p_3 = [1, 0], x_1 = [1.5, 3], \) then the MC simulation results of \( x_i(t + 1) \) (marked by blue points) and the theoretical values of \( D_i \) (as calculated via Eq. (3.12), and marked by blue dotted rectangles) are shown in Fig. 7.

![Fig. 7. MC simulation results of \( x_i(t+1) \) (blue points), (a): \( a(t) = 2 \), (b): \( a(t) = 0.5 \)](image)

It can be seen from Fig. 7 that the distribution of \( x_i(t + 1) \) approximately forms a rectangle, whose center is \( \frac{1}{3} \sum p \) (marked by the red dot in Fig. 7). In addition, the probability that \( x_i(t + 1) \)
appears next to $\frac{1}{3} \Sigma p$ is obviously larger than the probability of being far away from $\frac{1}{3} \Sigma p$, which coincides well with the conclusion that the $h_{ij}(u_j)$ is a single-peak function (please see the corollary 2.2). However, different from the MC simulation results of $x'_k$, the side-length of the rectangle envelope of the simulation results of $x_i(t + 1)$ (see the blue points as shown in Fig. 7) is evidently smaller than that of the theoretical value of $D_i$ (calculated by Eq. (3.12), and marked by the blue dotted rectangle as shown in Fig. 7), which indicates that the probability that $x_{ij}(t + 1)$ becomes close to its theoretical maximum (or minimum) value is remarkably smaller than that of $x_{kj}(t)$.

In order to illustrate this phenomenon more directly, we generate the curves of $h_{ij}(u_j)$ and $g_{kj}(u_j)$ in the same coordinate system by MC simulations as shown in Fig. 8 (a). Let $a(t) = 2, p_{1j} = p_{2j} = p_{3j} = 1, x_{ij} = 3$, then it can be calculated that both the domains of $h_{ij}(u_j)$ and $g_{kj}(u_j)$ are $[-5,7]$ (please see the corollary 1.1 and corollary 2.1). In addition, based on the numerical results of $h_{ij}(u_j)$ and $g_{kj}(u_j)$, the CDF of $x'_k(t)$ (marked as $G_{kj}(u_j)$) and the CDF of $x_{ij}(t + 1)$ (marked as $H_{ij}(u_j)$) are calculated via Trapezoid formula [58], and the curves of $G_{kj}(u_j)$ and $H_{ij}(u_j)$ are drawn in the same coordinate system as well, which can be found in Fig. 8 (b). It can be seen that when $u_j$ is close to -5, the value of $H_{ij}(u_j)$ is obviously smaller than that of $G_{kj}(u_j)$, and the value of $h_{ij}(u_j)$ is obviously smaller than that of $g_{kj}(u_j)$, which indicates that the probability that $x_i(t + 1)$ becomes close to its boundary is remarkably smaller than that of $x'_k$.

(a) 
(b) 

Fig. 8. (a): Curves of $h_{ij}(u_j)$ and $g_{kj}(u_j)$, (b): Curves of $H_{ij}(u_j)$ and $G_{kj}(u_j)$.

4. Analytical model of sampling distribution of $x_i(t + 1)$ without constant $x_i(t)$ assumption

In section 3, under both stagnation and constant $x_i(t)$ assumptions, we derived and analyzed the distribution that $x_i(t + 1)$ follows. Based on these works, in this section, only the stagnation assumption is considered to analyze the sampling distribution that $x_i(t + 1)$ follows, in which condition $x_i(t + 1)$ should be regarded as a stochastic process rather than a simple random variable. Although the exact analytical expression of the joint PDF of $x_i(t)$ is hard to derive, several basic characteristics of the distribution can be obtained. Firstly, we will discuss the shape of the PDF of $x_{ij}(t)$, and then derive the moments and central moments of $x_{ij}(t)$, which will serve as important tools to understand the distribution that $x_{ij}(t)$ follows. Finally, as an application of the results derived above, the order-1 and order-2 stability of GWO under the stagnation assumption is introduced and proved.
4.1 Characteristics of the PDF of $x_i(t)$

In this section, in order to better illustrate that the distribution of $x_{ij}(t)$ is related to iteration $t$, we mark the PDF of $x_{ij}(t + 1)$ as $h_{ij}(u_j, t)$ rather than $h_{ij}(u_j)$ (as marked in section 3).

The domain of $h_{ij}(u_j, t)$ is firstly discussed. Owing to the existence of absolute value sign of the domain of $h_{ij}(u_j)$ under constant $x_{ij}(t)$ assumption (please see corollary 2.1, Eqs. (3.4) and (3.9)), it is hard to derive the analytical expression of the domain of $h_{ij}(u_j, t)$ recursively. However, we could easily prove that:

**Proposition 3.1**

When $t \geq 2$, the domain of $h_{ij}(u_j, t)$ contains $\left(\frac{1}{3} \sum p_j - d_1 \prod_{\tau=1}^{t-2} a(\tau), \frac{1}{3} \sum p_j + d_1 \prod_{\tau=1}^{t-2} a(\tau)\right)$, where $d_1 = \max_{x_{ij}(1)} \big| x_{ij}(1) - \frac{1}{3} \sum p_j \big|$.

Proposition 3.1 shows that the domain of $h_{ij}(u_j, t)$ contains an open interval with center $\frac{1}{3} \sum p_j$ and length $2d_1 \prod_{\tau=1}^{t-2} a(\tau)$. It can be seen that when $t < \frac{T}{2}, a(t) > 1$, the length is an increasing function of iteration $t$, which means in the first half period of the GWO, the searching range of each agent increases.

The next conclusion is pertinent to the shape of the curve of $h_{ij}(u_j, t)$:

**Proposition 3.2**

$\forall t \geq 2, h_{ij}(u_j, t)$ is a single-peak function symmetrical about $u_j = \frac{1}{3} \sum p_j$.

According to the proposition 3.2, even if the assumption that $x_i(t)$ is constant is eliminated, the PDF of $x_{ij}(t + 1)$ is still a single-peak function symmetrical about $u_j = \frac{1}{3} \sum p_j$ (please see the corollary 2.2 for the conclusion under constant $x_i(t)$ assumption).

The propositions above can be verified by MC simulations. Here, the sample size is set to be $10^6$, and the number of frequency histogram segments is set to be 80. Let $p_{1j} = -1, p_{2j} = 1.5, p_{3j} = 2.5$ (i.e. $\frac{1}{3} \sum p_j = 1$) , $x_{ij}(1) \sim U[-4,4]$ (corresponding to the optimization problem with feasible solution $[-4,4])$, maximum iterations $T = 60$ (i.e. $a(t) = 2 \left(1 - \frac{t}{60}\right)$), and iteration $t = 2,5,10,20,30,50$. The corresponding obtained frequency histograms of $x_{ij}(t)$ are shown in Fig. 9 (a) – (f), respectively.
Fig. 9. Frequency histogram of $x_{ij}(t)$ under stagnation assumption, (a)-(f): $t = 2, 5, 10, 20, 30, 50$.

It can be seen from Fig. 9 that whatever the value of $t$ is, the curve of the PDF of $x_{ij}(t)$ is unimodal and symmetrical about $u_j = \frac{1}{3} \sum p_j = 1$. However, as iteration $t$ increases, the distribution of $x_{ij}(t)$ continuously becomes more close to $\frac{1}{3} \sum p_j$ (please see the variation of scales of X-axis from Fig. 9 (a) to Fig. 9 (f)), which seems to be contradictory with proposition 3.1. This is because the probability that $x_{ij}(t)$ is close to its bound of domain is rather small (please see Fig. 8 and related discussions), and the probability becomes even smaller after iterations. Thus, although the domain of $h_{ij}(u_j, t)$ expands during the first half period of the GWO, it is safe to assume that the distribution of $x_{ij}(t)$ becomes closer to its center $\frac{1}{3} \sum p_j$ during iterations.
The proposition 3.1 provides a preliminary study about the characteristics of \( h_{ij}(u_j, t) \). In order to better understand the distribution that \( x_{ij}(t) \) follows, in the next section, we will derive the recursive expression of the central moments of any positive integer order of \( x_{ij}(t) \).

4.2 Moments of \( x_{ij}(t) \)

It is known that the moment is a crucial characteristic of a random variable. In fact, if the moment of any order of a random variable is known, then the distribution of the random variable will be defined uniquely. As for the \( x_{ij}(t) \), although the PDF of \( x_{ij}(t) \) cannot be derived directly, the moments of any positive integer order of \( x_{ij}(t) \) can be calculated as follows.

Firstly, according to the proposition 3.2, \( \forall t, \mathbb{E}x_{ij}(t) = \frac{1}{3} \sum p_j \) holds. Here, we mark the \( r \)-order central moment of random variable \( X \) as \( \sigma^r X \), according to Eqs. (2.1) and (2.2), under the stagnation assumption, the \( r \)-order central moment of \( x_{ij}(t + 1) \) can be calculated as

\[
\sigma^r x_{ij}(t + 1) = \mathbb{E} \left( x_{ij}(t + 1) - \mathbb{E} x_{ij}(t + 1) \right)^r = \mathbb{E} \left( x_{ij}(t + 1) - \frac{1}{3} \sum p_j \right)^r, \tag{4.1}
\]

Since \( A_{kj} \sim U[-\alpha(t), \alpha(t)], C_{kj} \sim U[0, 2] \), for a positive integer \( r \), it can be calculated that

\[
\sigma^r x_{ij}(t + 1) = \frac{1}{3} \sum p_j \sum_{p=0}^{r} \sum_{q=0}^{2p} \binom{r}{2p} (\frac{r-2p}{2q}) \mathbb{E} \left( A_{1j}^p C_{1j}^{q} \left[ x_{ij}(t) \right]^{2p} A_{2j}^{2q} C_{2j}^{p} x_{ij}(t) - x_{ij}(t) \right)^{2q} C_{3j}^{p} \]

\[
= \frac{a(t)^r}{3^r} \sum_{p=0}^{r} \sum_{q=0}^{2p} \binom{r}{2p} (\frac{r-2p}{2q}) \mathbb{E} \left( \sum_{i=0}^{2q} C_{1j}^{q} \left[ x_{ij}(t) \right]^{2q} C_{3j}^{p} \right) \]

\[
= \frac{a(t)^r}{3^r} \sum_{p=0}^{r} \sum_{q=0}^{2p} \binom{r}{2p} (\frac{r-2p}{2q}) \mathbb{E} \left( \sum_{i=0}^{2q} C_{1j}^{q} \left[ x_{ij}(t) \right]^{2q} C_{3j}^{p} \right) \]

\[
= \frac{a(t)^r}{3^r} \sum_{p=0}^{r} \sum_{q=0}^{2p} \binom{r}{2p} (\frac{r-2p}{2q}) \mathbb{E} \left( \sum_{i=0}^{2q} C_{1j}^{q} \left[ x_{ij}(t) \right]^{2q} C_{3j}^{p} \right) \]

\[
= \frac{a(t)^r}{3^r} \sum_{p=0}^{r} \sum_{q=0}^{2p} \binom{r}{2p} (\frac{r-2p}{2q}) \mathbb{E} \left( \sum_{i=0}^{2q} C_{1j}^{q} \left[ x_{ij}(t) \right]^{2q} C_{3j}^{p} \right) \]

\[
= \frac{a(t)^r}{3^r} \sum_{p=0}^{r} \sum_{q=0}^{2p} \binom{r}{2p} (\frac{r-2p}{2q}) \mathbb{E} \left( \sum_{i=0}^{2q} C_{1j}^{q} \left[ x_{ij}(t) \right]^{2q} C_{3j}^{p} \right) \]

where in the first equality, the reason that the degree of each item in the right-hand side is even is that when \( r \) is odd, \( \mathbb{E} A_{kj}^r = 0 \) holds; the conclusion that \( \mathbb{E} A_{kj}^r = \frac{a(t)^r}{r+1} \) when \( r \) is even is applied to derive the right-hand side of the second equality; and the conclusion that \( \mathbb{E} C_{kj}^r = \frac{2^r}{r+1} \) is applied to
derive the right-hand side of the last equality. In addition, the expression \( \frac{r}{z} - p - q \) is marked as \( s \) for symmetry of the expressions above.

According to the Eq. (4.2), \( \sigma^r x_{ij}(t + 1) \) can be expressed as the sum of several items (corresponding to the five-fold sum \( \sum_{m=0}^{r} \sum_{p=0}^{r-2p} \sum_{q=0}^{r-2p} \sum_{i=0}^{2q} \sum_{l=0}^{2s} \mathcal{S}_{ij}(l, q, i, p, m) \), where each item can be expressed as the linear combination of \( \mathbb{E} x_{ij}^m(t)(m = 0, 1, \ldots, r) \) (corresponding to the coefficient \( \frac{(r^2)^r}{2(2p+1)(2q+1)(2s+1)} \sum_{i=0}^{2q} \sum_{l=0}^{2s} \mathcal{S}_{ij}(l, q, i, p, m) \)). In summary, \( \sigma^r x_{ij}(t + 1) \) can be expressed as the linear combination of \( \mathbb{E} x_{ij}^m(t)(m = 0, 1, \ldots, r) \).

In addition, \( \mathbb{E} x_{ij}^r(t) \) can be calculated as:

\[
\mathbb{E} x_{ij}^r(t) = \mathbb{E} \left( (x_{ij}(t) - \frac{1}{3} \sum p_j) + \frac{1}{3} \sum p_j \right)^r = \sum_{m=0}^{r} \binom{r}{m} \left( \frac{1}{3} \sum p_j \right)^{r-m} \mathbb{E} \left( x_{ij}(t) - \frac{1}{3} \sum p_j \right)^m
\]

(4.3)

According to Eq. (4.3), \( \mathbb{E} x_{ij}^r(t) \) can be expressed as the linear combination of \( \sigma^m x_{ij}(t)(m = 0, 1, \ldots, r) \), which means by substituting Eq. (4.3) into Eq. (4.2), \( \sigma^r x_{ij}(t + 1) \) can be expressed as the linear combination of \( \sigma^m x_{ij}(t)(m = 0, 1, \ldots, r) \) as well. Thus, if the expression of all \( m \)-order \( (m = 0, 1, \ldots, r) \) central moments of \( x_{ij}(t) \) has been derived, we can obtain a difference equation of \( \sigma^r x_{ij}(t) \) corresponding to \( t \):

\[
\sigma^r x_{ij}(t + 1) = f_r \sigma^r x_{ij}(t) + g_r(t)
\]

(4.4)

where \( f_r \) is a coefficient, and \( g_r(t) \) denotes the linear combination of \( \sigma^m x_{ij}(t)(m = 0, 1, \ldots, r - 1) \).

Eq. (4.4) provides a way of solving the central moment of arbitrary positive integer order of \( x_{ij}(t) \) via mathematical induction: from \( r = 1 \), based on the expressions of \( \sigma^m x_{ij}(t)(m = 0, 1, \ldots, r) \) that have been solved, according to Eq. (4.4), the difference equation that \( \sigma^{r+1} x_{ij}(t) \) satisfies corresponding to \( t \) is solved. It is worth noting that for any odd positive integer \( r \), \( \sigma^r x_{ij}(t) = 0 \) holds, thus the procedure of solving \( \sigma^r x_{ij}(t) \) \( (r_0 \) is even) can be further clarified in Fig. 10:

![Fig. 10 Pseudo-code of solving \( \sigma^r x_{ij}(t) \) \( (r_0 \) is even).](image-url)
As an example, the second order central moment (i.e. variance) of \( x_{ij}(t) \) (marked as \( \mathbb{D}x_{ij}(t) \)) can be solved as:

**Proposition 3.3**

\[
\mathbb{D}x_{ij}(t + 1) = \frac{a(t)^2}{9} \left( \mathbb{D}x_{ij}(t) + 4 \frac{4}{9} \sum p_j^2 - \frac{1}{9} \left( \sum p_j \right)^2 \right) \tag{4.5}
\]

In the rest part of this section, the characteristics of \( \mathbb{D}x_{ij}(t) \) will be further discussed, which could serve as tools for the stability analysis of the GWO to be carried out in the next section. For simplification, \( \mathbb{D}x_{ij}(t) \) is abbreviated as \( D_t \), \( \frac{a(t)^2}{9} \) is abbreviated as \( b_t \) (i.e. \( b_t = \frac{4}{9} \left( 1 - \frac{\varepsilon}{T} \right)^2 \)), and expression \( \frac{4}{9} \sum p_j^2 - \frac{1}{9} \left( \sum p_j \right)^2 \) is marked as \( p_0 \) (obviously \( p_0 = \frac{1}{9} \sum p_j^2 + \frac{1}{9} \sum (p_i - \overline{p})^2 \geq 0 \) holds), then Eq. (4.5) can be rewritten as:

\[
D_{t+1} = b_t(D_t + p_0) \tag{4.6}
\]

The Eq. (4.6) can be seen as a first-order dynamic system. We firstly suppose that \( b_t \) is constant (independent of \( t \), since \( \forall t = 1, \ldots, T \), \( 0 \leq b_t < 1 \) holds, this dynamic system holds a unique attractor \( D_0 = \frac{b_0 p_0}{1 - b_t} \), which means if \( b_t \) is constant, \( \lim_{t \to \infty} D_t = D_0 \). In addition, it is worth noting that this assumption is approximately satisfied when the value of \( T \) is large enough (in which case \( |b_{t+1} - b_t| = \frac{4(2T^2 - 2T - 1)}{9T^2} \) tends to 0 as \( T \) increases), thus it could be assumed that when \( T \to \infty, \lim_{t \to \infty} D_t = D_0 \) holds. The analysis above can be summarized as follows:

**Corollary 3.1**

When \( T \to \infty \), \( D_t - \frac{b_0 p_0}{1 - b_t} \leq \left( \frac{4}{9} \right)^{t-1} \left| D_1 - \frac{b_0 p_0}{1 - b_t} \right| \) holds.

According to the corollary 3.1, an upper bound of the distance between \( D_t \) and \( D_0 = \frac{b_0 p_0}{1 - b_t} \) can be obtained when \( T \to \infty \), by which we can deduce that when \( T \to \infty \), \( \lim_{t \to \infty} D_t = D_0 \) holds.

The conclusions above can be verified via MC simulations. Here, we set \( T = 50 \), and the values of \( p_{kj} \) are randomly chosen. For each group of \( p_{kj} \) chosen, the GWO under the stagnation assumption is run for \( 10^6 \) trials independently (total iteration \( T = 50 \)). If the value of \( x_{ij} \) obtained in the \( l \)th trial with iteration \( t \) is marked as \( x_{ij}(l, t) \), and \( \forall l, x_{ij}(l, 1) \) is chosen as a random variable following \( U[-4,4] \), then \( D_t \) can be approximated as the variance of \( x_{ij}(l, t), l = 1, \ldots, 10^6 \). The \( D_t \) obtained from MC simulations and \( D_0 = \frac{b_0 p_0}{1 - b_t} \) are compared as shown in Fig. 11.
It can be seen from Fig. 11 that with the increasing of \( t \), the value of \( D_t \) tends to \( D_0 \). In addition, for each \( T \), the MC simulation is carried out, and \( |D_T - D_0| \) is calculated and shown in Table II, from which we can see that the value of \( |D_T - D_0| \) is close to 0, and decreases with the increasing of \( T \), which verifies the conclusion that \( \lim_{T \to \infty} D_T = D_0 \).

| \( T \) | \( |D_T - D_0| \) |
|------|------------------|
| 20   | \( 2.3 \times 10^{-3} \) |
| 30   | \( 1.7 \times 10^{-3} \) |
| 50   | \( 4.6 \times 10^{-4} \) |
| 100  | \( 3.7 \times 10^{-5} \) |
| 200  | \( 7.1 \times 10^{-6} \) |
| 500  | \( 2.2 \times 10^{-6} \) |

By calculating and analyzing the moments of \( x_{ij}(t) \), the characteristics of the updating mechanism of the GWO can be further understood. In the part II of the paper, we will prove the global convergence of the GWO based on the results of the moments of \( x_{ij}(t) \) calculated above. In addition, as another example, in the next section, the stability analysis of the GWO under the stagnation assumption will be carried out.

4.3 Stability analysis of GWO under the stagnation assumption

As an application of the results of \( \mathbb{E}x_{ij}(t) \) and \( \mathbb{D}x_{ij}(t) \) derived in the section 4.2, the stability analysis of GWO under the stagnation assumption is proved and discussed. As introduced in section 1, as an important branch of the theoretical analysis of meta-heuristic algorithms, the stability analysis has been carried out deeply, aiming at studying the dynamics of search agents, and moreover, on what kind of conditions do the search agents converge to some certain position [59].

Taking the PSO as an example, early analysis often ignores the randomness and sets the random coefficients as constants (whose values are often set as their expectations) [60-61], by which the updating mechanism of PSO could be simplified into a deterministic difference equation, thus the stability analysis could be performed in the deterministic dynamic systems by using methods such as Lyapunov stability analysis [62]. Obviously, the assumption that all the random coefficients are constants is too strong to fit the actual situation well. Subsequently, the stability analysis performed on the PSO with randomness consideration is widely studied. When randomness presents, a popular substitute called order-1 stability is defined as [46]:

**Definition 1**: A meta-heuristic algorithm holds order-1 stability if for any search agent \( i \), \( \lim_{t \to \infty} \mathbb{E}x_{ij}(t) \) exists.

Obviously, in the concept of order-1 stability, the deterministic model is established by
replacing the random variable \( x_{ij}(t) \) with \( \mathbb{E}x_{ij}(t) \).

Despite the fruitful research achievements of the order-1 stability of the PSO [62-64], it has been pointed out that such order-1 stability is not enough to ensure convergence, and the following definition of order-2 stability should be taken into account [46,65]:

**Definition 2**: A meta-heuristic algorithm holds order-2 stability if for any search agent \( i \),

\[
\lim_{t \to \infty} \mathbb{D}x_{ij}(t) = 0.
\]

It is worth noting that till now, for simplicity, the stability analysis of the PSO is mostly carried out under the stagnation assumption, in which the global best (gbest) of the population and the personal best (pbest) of each search agent are set to be constant (as mentioned in section 3.1).

Based on the analysis above, as a direct corollary of the proposition 3.2, the stability analysis of GWO under the stagnation assumption is performed as follows:

**Proposition 3.4**

The GWO holds order-1 stability under the stagnation assumption.

and

**Proposition 3.5**

The GWO holds order-2 stability under the stagnation assumption.

In the last part of this section, the two propositions above are verified by numerical simulations. The simulation method is identical with that of obtaining \( D_t \) and \( D_0 \) in Fig. 11 in section 4.2, i.e., \( T = 50 \), four groups of \( p_{kj} \) are randomly chosen, and for each group of \( p_{kj} \) chosen, the GWO under stagnation assumption is run for \( 10^6 \) trials independently. Here, we mark the value of \( x_{ij} \) obtained in the \( h \)th trial with iteration \( t \) as \( x_{ij}(l, t) \) (in each trial the value of \( x_{ij}(l, 1) \) is randomly chosen within \([-4,4]\), corresponding to the optimization problem with feasible region \([-4,4]\)), then \( \mathbb{E}x_{ij}(t) \) could be approximated as the mean of \( x_{ij}(l, t) \) (i.e. \( 10^{-6}\sum_{l=1}^{10^6}x_{ij}(l, t) \)), and \( \mathbb{D}x_{ij}(t) \) could be approximated as the variance of \( x_{ij}(l, t) \) (i.e. \( \frac{1}{10^6-1}\sum_{l=1}^{10^6}(x_{ij}(l, t) - 10^{-6}\sum_{l=1}^{10^6}x_{ij}(m, t))^2 \)). It is worth noting that after the stagnation assumption, whatever the value of \( f(x_i(t)) \) is, the position of the best three search agents will not be updated, which means the information of objective function is no longer required during simulations. The simulation results of \( \mathbb{E}x_{ij}(t) \) and \( \mathbb{D}x_{ij}(t) \) are shown in Fig. 12 and Fig. 13 (in which the Y-axis is set as \( \ln \mathbb{D}x_{ij}(t) \) to better exhibit the decreasing speed of \( \mathbb{D}x_{ij}(t) \)), respectively.

It can be seen from Fig. 12 that \( \mathbb{E}x_{ij}(t) = \begin{cases} 0, & t = 1 \\ \frac{1}{2}\sum p_j, & \text{otherwise} \end{cases} \) which is a direct inference of \( x_{ij}(t) \sim U[-4,4] \) and proposition 3.2. In addition, it can be seen from Fig. 13 that when \( t \to T \), the value of \( \ln \mathbb{D}x_{ij}(t) \) is negative and decreases sharply, thus the conclusion \( \lim_{t \to \infty} \mathbb{D}x_{ij}(t) = 0 \) can be verified.
Fig. 12 Simulation results of $\mathbb{E}x_{ij}(t)$, (a): $\frac{1}{3}\sum p_j = -1.92$, (b): $\frac{1}{3}\sum p_j = -1.14$, (c): $\frac{1}{3}\sum p_j = 0.17$, (d): $\frac{1}{3}\sum p_j = -0.28$
**5. Conclusion**

In the part I of the paper, we establish a theoretical framework of the GWO and discuss several promising theoretical findings, containing the sampling distribution, the order-1 and order-2 stability, and the global convergence analysis. According to the stagnation assumption, that is, the positions of the best three search agents \( p_k(t)(k = 1, 2, 3) \) remain unchanged during iterations, the characteristics of the sampling distribution of the new solution \( x_i(t + 1) \) and the probabilistic stability of the GWO are elaborately addressed.

For simplification, the characteristics of the sampling distribution of the new solution \( x_i(t + 1) \) is firstly discussed under the assumption that the original solution before updating \( x_i(t) \) is constant, in which case \( x_i(t + 1) \) becomes a function of random variables \( A_{kj}, C_{kj}(k = 1, 2, 3) \). Although the exact analytical expression of \( x_{ij}(t + 1) \)'s PDF (marked as \( h_{ij}(u_j) \)) is hard to derive, it can be proved that \( h_{ij}(u_j) \) is a continuous single-peak function symmetrical about \( u_j = \frac{1}{3} \sum p_j \) with a bounded domain. In addition, the differentiability of \( h_{ij}(u_j) \) is carefully studied.

Regarding more realistic condition that \( x_i(t) \) is no longer regarded as a constant, the conclusion that the \( h_{ij}(u_j) \) is a single-peak function symmetrical about \( u_j = \frac{1}{3} \sum p_j \) still holds. In addition, the central moments of any positive integer order of \( x_{ij}(t) \) is analyzed, and the order-1 and order-2 stability of the GWO under stagnation assumption is also proved. All the above analysis results are verified by numerical simulations.

The corresponding results obtained in the part I will serve as the basis for the global convergence analysis of the GWO in the part II. At the same time, we also believe that the theoretical findings in the part I of the paper can be used as a fundamental tool to some extent to help researchers better understand the searching mechanism and particle trajectory of the GWO, and future research on the application, novel variants, and more in-depth theoretical findings of the GWO can be carried out based on our works.
Appendix

Proof of proposition 1.2

Mark $g_{kj}(u_j)$ as $g_{kj}(u_j, x_{ij}, p_k)$, then according to Eq. (2.2), $g_{kj}(u_j, x_{ij}, p_k) = g_{kj}(u_j, x_{ij}, p_k) = g_{kj}(u_j, 2p_k - x_{ij}, p_k) = g_{kj}(2p_k - u_j, x_{ij}, p_k)$, thus we only need to consider the conditions $p_k \geq 0, x_{ij} \geq p_k$, and $u_j \geq p_k$.

Let $D = \{(A_{kj}, C_k) | p_k + A_{kj} | C_k | p_k - x_{ij} | \leq u_j, -a(t) \leq A_{kj} \leq a(t), 0 \leq C_k \leq 2\}$, then according to Eq. (2.2), the cumulative distribution function (CDF) of $x'_{kj}$ can be derived as

$$G_{kj}(u_j) = \int_D \frac{1}{4a(t)} dA_t dC_i.$$  

When $x_{ij} > 2p_k$ and $u_j > a(t)(x_{ij} - 2p_k)$, region $D$ is shown in Fig. 14 (a) (marked by grey shading, similarly hereinafter). In this case, we have

$$G_{kj}(u_j) = \frac{1}{2} + \frac{1}{4a(t)} \frac{u_j}{p_k} \ln \frac{x_{ij}}{x_{ij} - 2p_k}.$$  

When $p_k \leq x_{ij} \leq 2p_k$ and $u_j > a(t)(2p_k - x_{ij})$, region $D$ is shown in Fig. 14 (b). In this case, we have

$$G_{kj}(u_j) = \frac{1}{2} + \frac{1}{4a(t)} \frac{u_j}{p_k} \ln \frac{x_{ij}}{x_{ij} - 2p_k} + a(t) \left(2 - \frac{x_{ij}}{p_k}\right).$$  

When $p_k \leq x_{ij} \leq 2p_k$ and $0 \leq u_j \leq a(t)(2p_k - x_{ij})$, region $D$ is shown in Fig. 14 (c). In this case, we have

$$G_{kj}(u_j) = \frac{1}{2} + \frac{1}{2a(t)} \left[\frac{u_j}{p_k} + \frac{u_j}{p_k} \ln \frac{a(t) x_{ij}}{u_j} + a(t) \left(2 - \frac{x_{ij}}{p_k}\right)\right].$$  

Expanding the equations above to other cases where $x_{ij}, p_k, u_j \in \mathbb{R}$, and after a simple arrangement with absolute value sign, we finally get Eqs. (3.4) - (3.6). □
Proof of corollary 2.1

According to the characteristics of convolution operator, the domain of $h_{ij}(u_j)$ is evident.

It is well known that the convolution operator holds evident smoothing effect, for instance, the convolution of a continuous function and an arbitrary function must be a continuous function as well [57], i.e., the curve of the arbitrary function selected (which might not be a continuous function) is smoothed. It is worth noting that when $m_{kj} > 0$, $g_{kj}(u_j)$ is continuous within $\mathbb{R}$, therefore we only need to consider the case that $m_{kj} \leq 0$ ($k = 1,2,3$). Since $g_{kj}(u_j)$ is continuous within $\mathbb{R}\backslash\{p_{kj}\}$ and $\lim_{u \to p_{kj}} g_{kj}(u_j) = +\infty$ holds (please see proposition 1.2), thus the continuity of $(g_{kj} * g_{kj})(u_j)$ can be proved only $(g_{kj} * g_{kj})(p_{kj} + p_{kj}) < +\infty$ holds.

When $m_{kj} \leq 0$ ($k = 1,2,3$), just let $|m_{kj}| \leq |m_{kj}|$, then $(g_{kj} * g_{kj})(p_{kj} + p_{kj})$ can be calculated as:

$$(g_{kj} * g_{kj})(p_{kj} + p_{kj}) = 2 \int_{p_{kj} - m_{kj}}^{p_{kj} + m_{kj}} g_{kj}(u_j) g_{kj}(p_{kj} + p_{kj} - u_j) du_j + 2 \int_{p_{kj} + m_{kj}}^{p_{kj} + 2m_{kj}} g_{kj}(u_j) g_{kj}(p_{kj} + p_{kj} - u_j) du_j$$

Fig. 14 Region $D$ (marked by grey shading), (a): $x_{ij} > 2p_{kj}$ and $u_j > a(t)(x_{ij} - 2p_{kj})$; (b): $x_{ij} > 2p_{kj}$ and $0 \leq u_j \leq a(t)(x_{ij} - 2p_{kj})$; (c): $p_{kj} \leq x_{ij} \leq 2p_{kj}$ and $u_j > a(t)(2p_{kj} - x_{ij})$; (d): $p_{kj} \leq x_{ij} \leq 2p_{kj}$ and $0 \leq u_j \leq a(t)(2p_{kj} - x_{ij})$. 
Proof of corollary 2.2

The first two conclusions are evident due to the characteristics of convolution operator. As for
the latter part, we first prove that 
\((g_{1j} * g_{2j})(u_j)\) is non-decreasing within \((-\infty, p_{1j} + p_{2j})\).

For the sake of simplification, let 
\(g_{10}(u_j) = g_{1j}(u_j + p_{1j}), g_{20}(u_j) = g_{2j}(u_j + p_{2j})\), then according to the corollary 1.1, both \(g_{10}(u_j), g_{20}(u_j)\) are non-decreasing on \(\mathbb{R}^\), and we only need to prove that 
\(h_0(u_j) = (g_{10} * g_{20})(u_j)\) is non-decreasing on \(\mathbb{R}^-\). Let \(u_j < u_j + \Delta u_j < 0\), we have

\[
\begin{align*}
&h_0(u_j + \Delta u_j) - h_0(u_j) \\
&= \int_{-\infty}^{u_j} g_{10}(v) \left( g_{20}(u_j + \Delta u_j - v) - g_{20}(u_j - v) \right) dv
\end{align*}
\]

Let \(c(v) = g_{20}(u_j + \Delta u_j - v) - g_{20}(u_j - v)\), then obviously 
\(c(2u_j + \Delta u_j - v) = -c(v)\),
and when \(v < u_j + \frac{1}{2} \Delta u_j\), \(c(v) \leq 0\). Thus

\[
\begin{align*}
&h_0(u_j + \Delta u_j) - h_0(u_j) \\
&= \int_{-\infty}^{u_j+\frac{1}{2} \Delta u_j} g_{10}(v) c(v) dv + \int_{u_j+\frac{1}{2} \Delta u_j}^{\infty} g_{10}(v) c(v) dv
\end{align*}
\]

\[
\begin{align*}
&= \int_{-\infty}^{u_j+\frac{1}{2} \Delta u_j} \left( g_{10}(v) - g_{10}(2u_j + \Delta u_j - v) \right) c(v) dv \\
&\geq 0
\end{align*}
\]

i.e. \(h_0(u_j)\) is non-decreasing on \(\mathbb{R}^-\), and consequently 
\((g_{1j} * g_{2j})(u_j)\) is non-decreasing within \((-\infty, p_{1j} + p_{2j})\). Similarly, \(h_{ij}(u_j)\) is non-decreasing within \((-\infty, \frac{1}{2} \sum_{k=1}^{2} p_{kj})\). □

Proof of corollary 2.3

Just assume that \(m_{kj} > 0\), in which case \(g_{3j}(u_j)\) is continuous and bounded within \(\mathbb{R}\). It
can be deduced from the corollary 2.1 that whatever the signs of \(m_{kj}(k = 1, 2)\) are, 
\((g_{1j} * g_{2j})(u_j)\) is always a continuous and bounded function within \(\mathbb{R}\). Mark \((g_{1j} * g_{2j})(u_j)\) as
\[ d(u_j) \], then we only need to prove that \( \frac{d}{du}(d \ast g_{3j})(u_j) \) exists and is continuous within \( \mathbb{R} \).

For the convenience of analysis, all subsection region endpoints of \( d(u_j) \) are set as (from left to right) \( [s_0, s_1, ..., s_q] \), and all subsection region endpoints of \( g_{3j}(u_j) \) are set as (from left to right) \( [t_0, t_1, ..., t_{q_2}] \). By introducing unit step function \( \epsilon(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \), the expressions of \( g_{1j}(u_j) \) and \( g_{2j}(u_j) \) can be rewritten as:

\[
\begin{align*}
\frac{d}{du}(u_j) &= \sum_{i=1}^{q_1} \epsilon(u_j - s_{i-1}) - \epsilon(u_j - s_i) \bigg| \frac{v_{1j}(u_j)}{v_{1j}(u_j)} \\
&= \sum_{i=1}^{q_1} \epsilon(u_j - t_{i-1}) - \epsilon(u_j - t_i) \bigg| \frac{w_{1j}(u_j)}{w_{1j}(u_j)} \\
\text{where } v_{1j}(u_j), w_{1j}(u_j) \text{ are continuous functions defined within } \mathbb{R}, \forall t_1 = 1, ..., q_1, \forall u_j \in [s_{i-1}, s_i], v_{1j}(u_j) = g_{1j}(u_j); \forall t_2 = 1, ..., q_2, \forall u_j \in [t_{i-1}, t_i], w_{1j}(u_j) = g_{2j}(u_j), \text{ and } \forall t_2 = 1, ..., q_1 - 1, v_{1j}(s_i) = v_{1j+1}(s_i) \text{ holds.}
\end{align*}
\]

The expression of \( (d \ast g_{3j})(u_j) \) can be expanded as:

\[
(d \ast g_{3j})(u_j) = \int_{-\infty}^{t_{\infty}} \sum_{i=1}^{q_1} \sum_{i=1}^{q_2} \epsilon(z - s_{i-1}) - \epsilon(z - s_i) \bigg| \frac{v_{1j}(z)}{v_{1j}(z)} \bigg| \frac{w_{1j}(u_j - z)}{w_{1j}(u_j - z)} dz + \\
= \sum_{i=1}^{q_1} \sum_{i=1}^{q_2} \int_{-\infty}^{t_{\infty}} \epsilon(x - s_{i-1}) - \epsilon(x - s_i) \bigg| \frac{v_{1j}(x)}{v_{1j}(x)} \bigg| \frac{w_{1j}(u_j - z)}{w_{1j}(u_j - z)} dz - \\
\int_{-\infty}^{t_{\infty}} \epsilon(z - s_{i-1}) - \epsilon(z - s_i) \bigg| \frac{v_{1j}(z)}{v_{1j}(z)} \frac{w_{1j}(u_j - z)}{w_{1j}(u_j - z)} dz - \\
\text{Thus}
\]

\[
\frac{d}{du}(d \ast g_{3j})(u_j)
\]

\[
= \sum_{i=1}^{q_1} \sum_{i=1}^{q_2} \int_{-\infty}^{t_{\infty}} \epsilon(z - s_{i-1}) - \epsilon(z - s_i) \bigg| \frac{v_{1j}(z)}{v_{1j}(z)} \bigg| \frac{w'_{1j}(u_j - z)}{w'_{1j}(u_j - z)} dz + \\
= \sum_{i=1}^{q_1} \sum_{i=1}^{q_2} \int_{-\infty}^{t_{\infty}} \epsilon(x - s_{i-1}) - \epsilon(x - s_i) \bigg| \frac{v_{1j}(x)}{v_{1j}(x)} \bigg| \frac{w'_{1j}(u_j - z)}{w'_{1j}(u_j - z)} dz - \\
\text{where } g'_{3j}(u_j) \text{ denotes the piecewise function in which each subsection } w_{1j}(u_j) \text{ is replaced by } w'_{1j}(u_j) = \frac{dv'_{1j}(u_j)}{du_j}, \text{ rather than the derivative of } g_{3j}(u_j) \text{ (in fact } g_{3j}(u_j) \text{ is non-derivable on its subsection region endpoints).}
\]

Since \( d(u_j) \) is continuous and bounded within \( \mathbb{R} \), the first item of \( \frac{d}{du}(d \ast g_{3j})(u_j) \) (i.e. \( (d \ast g'_{3j})(u_j) \)) is continuous within \( \mathbb{R} \) as well. Since \( g_{3j}(u_j) \) is bounded within \( \mathbb{R} \) and \( \forall t_2 = 1, ..., q_1 - 1, v_{1j}(s_i) = v_{1j+1}(s_i) \) holds, the second item of \( \frac{d}{du}(d \ast g_{3j})(u_j) \) is continuous within
\(\mathbb{R}\) as well. In summary, when there exists at least one positive integer among \(m_k (k = 1, 2, 3)\), \(h_{ij} (u_j)\) is differentiable within \(\mathbb{R}\).

In addition, since \(h_{ij} (u_j + \frac{1}{3} \sum p_j)\) is an even function, \(h'_{ij} (u_j + \frac{1}{3} \sum p_j)\) becomes an odd function. It is worth noting that \(h'_{ij} (u_j + \frac{1}{3} \sum p_j)\) is continuous within \(\mathbb{R}\), thus we have \(h'_{ij} (\frac{1}{3} \sum p_j) = 0\). □

Proof of proposition 3.1

Obviously, \(\forall t \geq 2\), the domain of \(h_{ij} (u_j, t)\) is symmetrical about \(\frac{1}{3} \sum p_j\). Mark the right endpoint of the domain of \(h_{ij} (u_j, t)\) as \(u(t)\), if \(x_{ij}(1)\) is fixed, then according to corollary 2.1, we have:

\[
\begin{align*}
u(2) - \frac{1}{3} \sum p_j &= \frac{1}{3} \sum n_j = \frac{a(1)}{3} [\sum_{k=1}^3 |p_k| + \sum_{k=1}^3 |x_{ij}(1) - p_k|] \\
&\geq \frac{a(1)}{3} \sum_{k=1}^3 |x_{ij}(1) - p_k| \\
&\geq a(1) \left| x_{ij}(1) - \frac{1}{3} \sum p_j \right|
\end{align*}
\]

Taking the fact that \(x_{ij}(1)\) is a random variable into consideration, we have \(u(2) - \frac{1}{3} \sum p_j \geq a(1) \max_{x_{ij}(1)} \left| x_{ij}(1) - \frac{1}{3} \sum p_j \right| = a(1)d_1\).

By repeating the derivations above, we can get:

\[
\begin{align*}
u(t) - \frac{1}{3} \sum p_j &\geq a(t - 1) \max_{x_{ij}(t-1)} \left| x_{ij}(t-1) - \frac{1}{3} \sum p_j \right| = a(t - 1) (u(t - 1) - \frac{1}{3} \sum p_j) \\
&\geq a(t - 1) \sum_{\tau=1}^{t-1} a(\tau) \geq \cdots \geq \sum_{\tau=1}^{t-1} a(\tau) \\
&\geq d_1 \sum_{\tau=1}^{t-1} a(\tau)
\end{align*}
\]

which means the domain of \(h_{ij} (u_j, t)\) contains \(\left( \frac{1}{3} \sum p_j - d_1 \sum_{\tau=1}^{t-1} a(\tau) \right) \frac{1}{3} \sum p_j + d_1 \sum_{\tau=1}^{t-1} a(\tau)\). □

Proof of proposition 3.2

Eq. (2.2) shows that \(x'_{k_j} (t) = p_{k_j} + A_{k_j} [C_{k_j} |p_{k_j} - x_{ij}(t)|]\), where \(A_{k_j} \sim U[-a, a]\), i.e. the distribution that \(A_{k_j}\) follows is symmetrical about \(u_j = 0\). Since \(A_{k_j}\) is independent of \([C_{k_j} p_{k_j} - x_{ij}(t)]\), whatever distribution \(x_{ij}(t)\) follows, the distribution that \(x'_{k_j}(t)\) follows must be symmetrical about \(\frac{1}{3} \sum p_j\), which means the curve of \(h_{ij} (u_j, t)\) is symmetrical about \(u_j = \frac{1}{3} \sum p_j\). Next, we prove that \(h_{ij} (u_j, t)\) is non-decreasing within \((-\infty, \frac{1}{3} \sum p_j)\) via mathematical induction.

(1) when \(t = 1\), according to the procedure of the GWO, \(x_{ij}(1)\) follows a uniform distribution within the feasible region, which makes the conclusion evident.
(2) if the conclusion is correct when \( t = t_0 \), i.e. \( h_{ij}(u_j, t_0) \) is non-decreasing within \((-\infty, \frac{1}{3}\sum p_j)\), then according to the corollary 2.2, as the sum of two random variables \(-x_{ij}(t_0)\) and \( C_{kj}p_{kj} \) that both follow single-peak symmetrical distributions, \( C_{kj}p_{kj} - x_{ij}(t_0) \) follows a single-peak symmetrical distribution as well. Since \( A_{kj} \) is independent of \( C_{kj}p_{kj} - x_{ij}(t_0) \), and the distribution that \( A_{kj} \) follows is symmetrical about \( u_j = 0 \), it can be seen that \( A_{kj}|C_{kj}p_{kj} - x_{ij}(t_0)\) and \( A_{kj}(C_{kj}p_{kj} - x_{ij}(t_0)) \) follow the same distribution, in which case we only need to prove that the PDF of \( A_{kj}(C_{kj}p_{kj} - x_{ij}(t_0)) \) is non-increasing within \((0, +\infty)\).

Let the PDF of \( A_{kj}(C_{kj}p_{kj} - x_{ij}(t_0)) \) be \( \varphi(u_j) \), by symmetry, just suppose that the symmetrical axis of the curve of \( \varphi(u_j) \) is on the left of Y-axis, which means \( \varphi(u_j) \) is non-increasing within \((0, +\infty)\); mark the PDF of \( A_{kj} \) as \( \theta(u_j) = \frac{1}{2a}(\varepsilon(u_j + a) - \varepsilon(u_j - a)) \), the PDF and CDF of \( A_{kj}(C_{kj}p_{kj} - x_{ij}(t_0)) \) as \( \delta(u_j) \) and \( \Delta(u_j) \) respectively, then the expression of \( \Delta(u_j) \) can be derived as:

\[
\Delta(u_j) = \int_{uv\leq u_j} \varphi(v)\theta(w)dw = \frac{1}{2} + 2\int_0^a \frac{1}{2a}dw \int_0^u \varphi(v)dv.
\]

Thus \( \delta(u_j) = \frac{d\Delta(u_j)}{du_j} = \frac{1}{a} \int_0^a w\varphi\left(\frac{u_j}{w}\right)dw \) holds.

Let \( \Delta u_j > 0 \), then \( \delta(u_j + \Delta u_j) - \delta(u_j) = \frac{1}{a} \int_0^a w\left[\varphi\left(\frac{u_j + \Delta u_j}{w}\right) - \varphi\left(\frac{u_j}{w}\right)\right]dw \). Since \( \varphi(u_j) \)

is non-increasing within \((0, +\infty)\), \( \forall w > 0, u_j > 0, \varphi\left(\frac{u_j + \Delta u_j}{w}\right) - \varphi\left(\frac{u_j}{w}\right) \leq 0 \) holds, thus \( \delta(u_j + \Delta u_j) - \delta(u_j) \leq 0 \) holds. Owing to the arbitrariness of positive numbers \( u_j, \Delta u_j \), we can deduce that \( \delta(u) \) is non-increasing within \((0, +\infty)\).

Finally, according to the corollary 2.2, we can prove that the PDF of \( x_{ij}(t_0 + 1) = \frac{1}{3}\sum p_j + \frac{1}{3}\sum_{k=1}^3 A_{kj}|C_{kj}p_{kj} - x_{ij}(t_0)| \) is non-decreasing within \((-\infty, \frac{1}{3}\sum p_j)\), which means \( \forall t, h_{ij}(u_j, t) \)

is non-decreasing within \((-\infty, \frac{1}{3}\sum p_j)\). □

**Proof of proposition 3.3**

According to Eq. (4.2):

\[
\mathbb{D} x_{ij}(t + 1) = \frac{1}{9} \mathbb{E}\left(\sum_{k=1}^3 A_{kj}|C_{kj}p_{kj} - x_{ij}(t)|\right)^2 \leq \frac{1}{6} \sum_{k=1}^3 \mathbb{E}A_{kj}^2 \mathbb{E}\left(C_{kj}p_{kj} - x_{ij}(t)\right)^2
\]

\[
= \frac{a(t)^2}{27} \sum_{k=1}^3 \left(\mathbb{E}x_{ij}^2(t) - 2p_{kj}\mathbb{E}x_{ij}(t) + \frac{4}{3}p_{kj}^2\right) = \frac{a(t)^2}{9} \left(\mathbb{E}x_{ij}^2(t) - \frac{2}{3}(\sum p_j)\mathbb{E}x_{ij}(t) + \frac{4}{3}\sum p_j^2\right)
\]

According to Eq. (4.3), \( \mathbb{D} x_{ij}(t) = \mathbb{E}\left(x_{ij}(t) - \frac{1}{3}\sum p_j\right)^2 = \mathbb{E}x_{ij}^2(t) - \frac{2}{3}(\sum p_j)\mathbb{E}x_{ij}(t) + \frac{4}{3}\sum p_j^2\).
\[ \frac{1}{9}(\Sigma p_i)^2. \]

According to the equations above, we can derive that \( \mathbb{D}x_{ij}(t + 1) = \frac{a(t)^2}{9} \left( \mathbb{D}x_{ij}(t) + \frac{4}{9} \Sigma p_j^2 - \frac{1}{9}(\Sigma p_j)^2 \right) \) holds. □

**Proof of corollary 3.1**

The following estimation of \( |D_{t+1} - \frac{b_{t+1}p_0}{1-b_{t+1}}| \) holds:

\[
\begin{align*}
|D_{t+1} - \frac{b_{t+1}p_0}{1-b_{t+1}}| & \leq |D_{t+1} - \frac{b_t p_0}{1-b_t}| + |\frac{b_t p_0}{1-b_t} - \frac{b_{t+1}p_0}{1-b_{t+1}}| = b_t |D_t - \frac{b_t p_0}{1-b_t}| + p_0 \frac{b_t - b_{t+1}}{(1-b_t)(1-b_{t+1})} \\
& \leq \frac{4}{9} |D_t - \frac{b_t p_0}{1-b_t}| + \frac{36p_0^2 2t+1}{25t^2}
\end{align*}
\]

where the second inequality is derived due to the fact that \( b_t \leq \frac{4}{9} \). Obviously \( \lim_{T \to \infty} \frac{36p_0^2 2t+1}{25t^2} = 0 \), thus when \( T \to \infty \), \( |D_t - \frac{b_t p_0}{1-b_t}| \leq \frac{4}{9} |D_{t-1} - \frac{b_{t-1}p_0}{1-b_{t-1}}| \leq \cdots \leq \left( \frac{4}{9} \right)^{t-1} |D_1 - \frac{b_1 p_0}{1-b_1}| \to 0 \) holds. □

**Proof of proposition 3.5**

In the GWO, we only need to prove that \( \lim_{T \to \infty} \mathbb{D}x_{ij}(T) = 0 \). The first-order dynamical system described in Eq. (4.6) can be solved as:

\[
\mathbb{D}x_{ij}(t) = \mathbb{D}x_{ij}(1) \prod_{r=1}^{t-1} b_r + p_0 b_{t-1} \left( 1 + \sum_{r=0}^{t-2} \prod_{r=1}^{t-2} b_r \right), t > 1
\]

Thus,

\[
\mathbb{D}x_{ij}(T) = \mathbb{D}x_{ij}(1) \prod_{r=1}^{T-1} b_r + p_0 b_{T-1} \left( 1 + \sum_{r=0}^{T-2} \prod_{r=1}^{T-2} b_r \right) < \mathbb{D}x_{ij}(1) b_{T-1} + p_0 b_{T-1} \left( 1 + \sum_{r=0}^{T-2} 1 \right)
\]

\[
= \frac{4}{9T^2} \left( \mathbb{D}x_{ij}(1) + p_0(T - 1) \right)
\]

Obviously \( \mathbb{D}x_{ij}(1) < +\infty \), thus \( \lim_{T \to \infty} \mathbb{D}x_{ij}(T) \leq \lim_{T \to \infty} \frac{4}{9T^2} \left( \mathbb{D}x_{ij}(1) + p_0(T - 1) \right) = 0 \), which means \( \lim_{T \to \infty} \mathbb{D}x_{ij}(t) = 0 \) holds. □

**Reference**

[1] H. Zang, S. Zhang, K. Hapeshi, A review of nature-inspired algorithms, Journal of Bionic Engineering, 7 (2010) S232-S237, https://doi.org/10.1016/S1672-6529(09)60240-7.

[2] P. A. Vikhar, Evolutionary algorithms: A critical review and its future prospects, 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC). IEEE, 2016, pp. 261-265, https://doi.org/10.1109/ICGTSPICC.2016.7955308.

[3] S. Boyd, L. Vandenberghe, Convex Optimization, 7th ed., Cambridge University Press,
Cambridge, 2004.

[4] K. Sörensen, Metaheuristics—the metaphor exposed, International Transactions in Operational Research, 22.1 (2013) 3-18, https://doi.org/10.1111/itor.12001.

[5] A. A. Heidari, S. Mirjalili, H. Faris, et al., Harris hawks optimization: Algorithm and applications, Future Generation Computer Systems, 97 (2019) 849-872. https://doi.org/10.1016/j.future.2019.02.028.

[6] J. H. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, MIT press, 1992.

[7] J. R. Koza, Genetic programming: on the programming of computers by means of natural selection, MIT press, 1992.

[8] D. B. Fogel, Artificial intelligence through simulated evolution, Wiley-IEEE Press, 1998.

[9] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by simulated annealing, science, 220 (1983) 671-680, https://doi.org/10.1126/science.220.4598.671.

[10] O. K. Erol, I. Eksin, A new optimization method: big bang–big crunch, Advances in Engineering Software, 37.2 (2006) 106-111, https://doi.org/10.1016/j.advengsoft.2005.04.005.

[11] R. A. Formato, Central Force Optimization: A New Nature Inspired Computational Framework for Multidimensional Search and Optimization, in: N. Krasnogor, G. Nicosia, M. Pavone, D. Pelta, et al., Nature Inspired Cooperative Strategies for Optimization (NICSO 2007). Studies in Computational Intelligence, vol 129, Springer, Berlin, Heidelberg, 2008, pp. 221-238, https://doi.org/10.1007/978-3-540-78987-1_21.

[12] F. Glover, M. Laguna, Tabu search, Handbook of combinatorial optimization, Springer, Boston, MA, 1998, pp. 2093-2229, https://doi.org/10.1007/978-1-4613-0303-9_33.

[13] R. V. Rao, V. J. Savsani, D. P. Vakharia, Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, Computer-Aided Design, 43.3 (2011) 303-315, https://doi.org/10.1016/j.cad.2010.12.015.

[14] Z. W. Geem, J. H. Kim, G. V. Loganathan, A new heuristic optimization algorithm: harmony search, simulation, 76.2 (2001) 60-68, https://doi.org/10.1177/003754970107600201.

[15] M. Dorigo, M. Birattari, T. Stutzle, Ant colony optimization, IEEE computational intelligence magazine, 1.4 (2006) 28-39, https://doi.org/10.1109/MCI.2006.329691.

[16] J. Kennedy, R. Eberhart, Particle swarm optimization, Proceedings of ICNN'95 international conference on neural networks, IEEE, 1995, pp. 1942-1948, https://doi.org/10.1109/ICNN.1995.488968.

[17] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, Journal of global optimization, 39 (2007) 459-471, https://doi.org/10.1007/s10898-007-9149-x.

[18] X. S. Yang S. Deb, Cuckoo search via Lévy flights, 2009 World congress on nature & biologically inspired computing (NaBIC), IEEE, 2009, pp. 210-214, https://doi.org/10.1109/NaBIC.2009.5393690.

[19] A. Chakraborty A, A. K. Kar, Swarm intelligence: A review of algorithms, in: S. Patnaik, X. S. Yang, K. Nakamatsu, Nature-Inspired Computing and Optimization. Modeling and Optimization in Science and Technologies, vol 10, Springer, Cham, 2017, pp. 475-494, https://doi.org/10.1007/978-3-319-50920-4_19.

[20] X. S. Yang, Nature-inspired metaheuristic algorithms, Luniver press, 2010.

[21] S. Salcedo-Sanz, Modern meta-heuristics based on nonlinear physics processes: A review of
models and design procedures, Physics Reports, 655 (2016) 1-70, https://doi.org/10.1016/j.physrep.2016.08.001.

[22] S. Mirjalili, S. M. Mirjalili, A. Lewis, Grey wolf optimizer, Advances in engineering software, 69 (2014) 46-61, https://doi.org/10.1016/j.advengsoft.2013.12.007.

[23] L. D. Mech, Alpha status, dominance, and division of labor in wolf packs, Canadian Journal of Zoology, 77.8 (1999) 1196-1203, https://doi.org/10.1139/z99-099.

[24] T. Jayabarathi, T. Raghunathan, B. R. Adarsh, et al., Economic dispatch using hybrid grey wolf optimizer, Energy, 111 (2016) 630-641, https://doi.org/10.1016/j.energy.2016.05.105.

[25] H. Wang, L. Shi, Y. Ni, Complementary and coordinated optimal dispatch solution for hybrid thermal-wind-PV power grid with energy storage, 8th Renewable Power Generation Conference (RPG 2019). IEEE, 2019, pp. 1-8, https://doi.org/10.1049/cp.2019.0544.

[26] F. Zhai, L. Shi, Solution proposal to the unit commitment problem incorporating manifold uncertainties, IET Generation, transmission & distribution, 14.21 (2020) 4763-4774, https://doi.org/10.1049/iet-gtd.2020.0805.

[27] S. Mirjalili, How effective is the Grey Wolf optimizer in training multi-layer perceptrons, Applied Intelligence, 43 (2015) 150-161, https://doi.org/10.1007/s10489-014-0645-7.

[28] M. R. Mosavi, M. Khishe, A. Ghamgosar. Classification of sonar data set using neural network trained by Gray wolf optimization, Neural Network World, 26.4 (2016) 393-415, https://doi.org/10.14311/NNW.2016.26.023.

[29] S. Zhang, Y. Zhou, Grey wolf optimizer based on Powell local optimization method for clustering analysis, Discrete Dynamics in Nature and Society, 2015 (2015) 1-17, https://doi.org/10.1155/2015/481360.

[30] V. Kumar, J. K. Chhabra, D. Kumar. Grey wolf algorithm-based clustering technique, Journal of Intelligent Systems, 26.1 (2017) 153-168, https://doi.org/10.1515/jisys-2014-0137.

[31] P. W. Tsai, T. T. Nguyen, T. K. Dao, Robot Path Planning Optimization Based on Multiobjective Grey Wolf Optimizer, in: J. S. Pan, J. W. Lin, C. H. Wang, X. Jiang, Genetic and Evolutionary Computing. ICGEC 2016. Advances in Intelligent Systems and Computing, vol 536. Springer, Cham, 2017, pp. 166-173, https://doi.org/10.1007/978-3-319-48490-7_20.

[32] C. Qu, W. Gai, M. Zhong, et al. A novel reinforcement learning based grey wolf optimizer algorithm for unmanned aerial vehicles (UAVs) path planning, Applied Soft Computing, 89 (2020) 106099. https://doi.org/10.1016/j.asoc.2020.106099.

[33] H. Faris, I. Aljarah, M. A. Al-Betar, et al., Grey wolf optimizer: a review of recent variants and applications, Neural Computing & Applications, 30 (2018) 413-435, https://doi.org/10.1007/s00521-017-3272-5.

[34] G. Negi, A. Kumar, S. Pant, et al., GWO: a review and applications, International Journal of System Assurance Engineering and Management, 12 (2021) 1-8, https://doi.org/10.1007/s13198-020-00995-8.

[35] N. M. Hatta, A. M. Zain, R. Sallehuddin, et al., Recent studies on optimisation method of Grey Wolf Optimiser (GWO): a review (2014–2017), Artificial Intelligence Review, 52 (2019) 2651-2683, https://doi.org/10.1007/s10462-018-9634-2.

[36] K. Avadh, P. K. Singh, Empirical Study of Grey Wolf Optimizer, in: M. Pant, K. Deep, J. Bansal, A. Nagar, K. Das, Proceedings of Fifth International Conference on Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing, vol 436. Springer, Singapore, 2016, pp. 1037-1049. https://doi.org/10.1007/978-981-10-0448-3_87.
[37] L. Rodriguez, O. Castillo, M. García, J. Soria, A Comparative Study of the Grey Wolf Optimizer and Firefly Algorithm in Mathematical Benchmark Functions of the CEC 15 Competition, in: P. Melin, O. Castillo, J. Kacprzyk, Recent Advances of Hybrid Intelligent Systems Based on Soft Computing. Studies in Computational Intelligence, vol 915. Springer, Cham, 2021, pp. 163-174, https://doi.org/10.1007/978-3-030-58728-4_9.

[38] L. I. Wong, M. H. Sulaiman, M. R. Mohamed, et al., Grey Wolf Optimizer for solving economic dispatch problems, IEEE International Conference on Power & Energy. IEEE, 2014, pp. 150-154, https://doi.org/10.1109/PECON.2014.7062431.

[39] V. K. Kamboj, S. K. Bath, J. S. Dhillon, Solution of non-convex economic load dispatch problem using Grey Wolf Optimizer, Neural Computing and Applications, 27 (2016) 1301-1316, https://doi.org/10.1007/s00521-015-1934-8.

[40] L. Rodriguez, O. Castillo, J. Soria, Grey wolf optimizer with dynamic adaptation of parameters using fuzzy logic, 2016 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2016, pp. 3116-3123, https://doi.org/10.1109/CEC.2016.7744183.

[41] M. Narayana, H. Nenavath, S. Chavan, et al., Intelligent visual object tracking with particle filter based on Modified Grey Wolf Optimizer, Optik, 193 (2019) 162913, https://doi.org/10.1016/j.ijleo.2019.06.013.

[42] A. A. Heidari, P. Pahlavani, An efficient modified grey wolf optimizer with Lévy flight for optimization tasks, Applied Soft Computing, 60 (2017) 115-134, https://doi.org/10.1016/j.asoc.2017.06.044.

[43] B. F. Azar, H. Veladi, F. Raeesi, et al., Control of the nonlinear building using an optimum inverse TSK model of MR damper based on modified grey wolf optimizer, Engineering Structures, 214 (2020) 110657, https://doi.org/10.1016/j.engstruct.2020.110657.

[44] K. Dudani, A. R. Chudasama, V. Sreeram, Partial discharge detection in transformer using adaptive grey wolf optimizer based acoustic emission technique, Cogent Engineering, 3.1 (2016) 1256083, https://doi.org/10.1080/23311916.2016.1256083.

[45] B. Yang, X. Zhang, T. Yu, et al., Grouped grey wolf optimizer for maximum power point tracking of doubly-fed induction generator based wind turbine, Energy Conversion and Management, 133 (2017) 427-443, https://doi.org/10.1016/j.enconman.2016.10.062.

[46] R. Poli, Mean and Variance of the Sampling Distribution of Particle Swarm Optimizers During Stagnation, IEEE Transactions on Evolutionary Computation, 13.4 (2009) 712-721, https://doi.org/10.1109/TEVC.2008.2011744.

[47] V. Kadirkamanathan, K. Selvarajah, P. J. Fleming, Stability analysis of the particle dynamics in particle swarm optimizer, IEEE Transactions on Evolutionary Computation, 10.3 (2006) 245-255, https://doi.org/10.1109/TEVC.2005.857077.

[48] H. P. Schwefel, Evolutionsstrategien für die numerische Optimierung, in: Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie. Interdisciplinary Systems Research / Interdisziplinäre Systemforschung (Analyse — Formalisierung — Simulation / Analysis — Modeling — Simulation). Birkhäuser, Basel, 1977, pp. 123-176, https://doi.org/10.1007/978-3-0348-5927-1_5.

[49] T. Back, Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms, Oxford university press, 1996.

[50] J. Jägersküpper, Algorithmic analysis of a basic evolutionary algorithm for continuous optimization, Theoretical Computer Science, 379.3 (2007) 329-347,
[51] X. Yao, Y. Liu, G. Lin, Evolutionary programming made faster, IEEE Transactions on Evolutionary Computation, 3.2 (1999) 82-102, https://doi.org/10.1109/4235.771163.

[52] K. A. Juste, H. Kita, E. Tanaka, J. Hasegawa, An evolutionary programming solution to the unit commitment problem, IEEE Transactions on Power Systems, 14.4 (1999) pp. 1452-1459, https://doi.org/10.1109/59.801925.

[53] H. Izadinia, M. M. Ebadzadeh, Adaptive Quantum-inspired Evolution Strategy, 2012 IEEE Congress on Evolutionary Computation, IEEE, 2012, pp. 1-8, https://doi.org/10.1109/CEC.2012.6256433.

[54] K. L. Chung, K. Zhong, A course in probability theory, Academic press, 2001.

[55] Weisstein, W. Eric, polylogarithm, https://mathworld.wolfram.com/Polylogarithm.html.

[56] P. Jiménez-Rodríguez, G. A. Muñoz-Fernández, E. Sáez-Maestro, et al., Algebraic genericity and the differentiability of the convolution, Journal of Approximation Theory, 241 (2019) 86-106, https://doi.org/10.1016/j.jat.2019.01.002.

[57] P. Jiménez-Rodríguez, S. Maghsoudi, G. A. Muñoz-Fernández, Convolution functions that are nowhere differentiable, Journal of Mathematical Analysis & Applications, 413.2 (2014) 609-615, https://doi.org/10.1016/j.jmaa.2013.12.008.

[58] J. H. Mathews, K. D. Fink, Numerical Methods Using MATLAB, 3rd ed., Pearson Education, Inc., 2001.

[59] Q. Liu, Order-2 stability analysis of particle swarm optimization, Evolutionary computation, 23.2 (2015) 187-216, https://doi.org/10.1116/EVCO_a_00129.

[60] F. Bergh, A. P. Engelbrecht, A study of particle swarm optimization particle trajectories, Information Sciences, 176.8 (2006), 937-971, https://doi.org/10.1016/j.ins.2005.02.003.

[61] M. Clerc, J. Kennedy, The particle swarm - explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation, 6.1 (2002), 58-73, https://doi.org/10.1109/4235.985692.

[62] V. Kadirkamanathan, K. Selvarajah, P. J. Fleming, Stability analysis of the particle dynamics in particle swarm optimizer, IEEE Transactions on Evolutionary Computation, 10.3 (2006) 245-255, https://doi.org/10.1109/TEVC.2005.857077.

[63] J. L. Fernández Martínez, E. G. Gonzalo, The generalized PSO: a new door to PSO evolution, Journal of Artificial Evolution & Applications, 2008 (2008) 1-15, https://doi.org/10.1155/2008/861275.

[64] V. Gazi, Stochastic stability analysis of the particle dynamics in the PSO algorithm, 2012 IEEE International Symposium on Intelligent Control, IEEE, 2012, pp. 708-713, https://doi.org/10.1109/ISIC.2012.6398264.

[65] M. Jiang, Y. P. Luo, S. Y. Yang, Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm. Information Processing Letters, 102.1 (2007) 8–16, https://doi.org/10.1016/j.ipl.2006.10.005.