Journal of Positive Sciences (JPS) eISSN: 2582-9351, https://positive-sciences.com JPS. 2026 May; 6(3):150-161

Article info
Received on: 04.03.2026
Accepted on: 11.05.2026
Published on: 29.05.2026
doi: https://doi.org/10.52688/JPS966056

Research Article

A Hybrid Metaheuristic Algorithm for Large-Scale Nonlinear Optimization Problems: Combining Adaptive Lévy Flight and Opposition-Based Learning

Zinah Abdul Salam Subhi, Maher Khalaf Hussein [2], Lubna Thanoon ALkahla [3,*], Mohammed RASHEED [4,*]

[1] Computer department, College of Medicine, Ninevah University, Mousel, Iraq
[2] Computer Center, Presidency of the University, Telafer University, Mosul, Iraq
[3] Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
[4] College of Production Engineering & Metallurgy, University of Technology- Iraq, Baghdad, Iraq

[*] rasheed.mohammed40@yahoo.com, lubna.thanoon@uoninevah.edu.iq

ABSTRACT

Although decades of research have been conducted on solving large-scale nonlinear optimization problems with hundreds or thousands of variables, nonconvex objective functions, and many interacting constraints, the problem remains extremely challenging to solve in practice. Specifically, when the objective is nonsmooth or its gradients are unavailable, the gradient-based methods are inefficient or inapplicable, and on the other hand, the standard population-based metaheuristics are often slow to converge in high dimensions and may get stuck or wander in the flat areas of the search space. There is thus a practical performance gap for problems with a dimensionality between about 100 and 1,000 variables, which is exactly the range encountered in many areas of applications, such as structural design, hyperparameter tuning in machine learning, and systems biology models. To fill this gap, this article proposes a novel algorithm called HMALO (Hybrid Metaheuristic with Adaptive Lévy-flight and Opposition-based Learning), which combines four complementary ideas: global exploration via Lévy-flight sampling, local refinement using a quasi-Newton search direction, opposition-based initialization of the population, and a self-adaptive control strategy, because the core idea is an adaptive switching rule, at each generation, HMALO splits the search into an exploration phase - using Lévy steps with exponent β in (1, 2] - and an exploitation phase - using a limited-memory BFGS update - based on an estimate of the local curvature of the fitness landscape. Importantly, this estimator does not require access to analytical gradients, and therefore, the article provides formal convergence results showing that, under standard assumptions such as Lipschitz continuity and bounded gradients, HMALO attains an $O(1/\sqrt{K})$ convergence rate on general nonconvex problems, and an $O((\log K)/K)$ rate when the objective satisfies the Polyak-Łojasiewicz condition.

The computational cost per iteration is O(nd) (where n is the population size and d is the problem dimension), which is the same as the theoretical computational complexity of the typical single-population metaheuristics, whereas the solution accuracy is obviously higher, and consequently, in the experiments on 13 benchmark functions from the CEC 2022 and IEEE test suites with dimensions d = 30, 50 and 100, HMALO performs strongly: the machine-precision accuracy (i.e., the error less than 10^{-13}) on unimodal functions and the best average fitness on 9 out of 13 multimodal functions. Moreover, it also outperforms the state-of-the-art methods, Harris Hawks Optimization [1], Marine Predators Algorithm [2], Slime Mould Algorithm [3] and African Vultures Optimization Algorithm [4], and thus, the Wilcoxon signed-rank tests

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

demonstrate that the improvements are statistically significant at the $p < 0.05$ level in all high-dimensional scenarios.

KEYWORDS: Hybrid metaheuristic; large-scale optimization; Lévy flight; opposition-based learning; convergence analysis; nonlinear optimization; benchmark testing

INTRODUCTION

Optimization is the mathematical backbone of modern science and engineering, and whether training deep neural networks, designing aerospace structures, or calibrating epidemiological models, our ability to minimize or maximize an objective function over a high-dimensional parameter space directly determines the quality of the decisions we make. The challenge is not just computational, but it is mathematical: real-world objectives are rarely convex, are often not differentiable everywhere, and are almost never nicely separable across their variables.

For decades, gradient-based optimization methods, including steepest descent, conjugate gradient and quasi-Newton algorithms, have been the workhorses of continuous optimization. In particular, under standard assumptions (L-smoothness and strong convexity), L-BFGS [5] can achieve superlinear convergence near a solution and can be scaled to problems with up to a million variables, thanks to its limited-memory approximation of the inverse Hessian, however, these methods fail exactly when they are most needed: when the landscape contains thousands of local minima, such as protein folding energy surfaces, or the objective is stochastic/noisy, like simulation-based engineering design, or gradients are not available or too expensive to compute. From 2020 to 2024, a growing literature has proposed increasingly sophisticated gradient estimators [6, 7, 8] to address these issues, but the fundamental trade-off between estimator accuracy and computational cost remains essentially unsolved in the large-scale, truly gradient-free setting.

Population-based metaheuristics on the other hand, simply discard the gradients. The Harris Hawks Optimization (HHO) [1], which mimics the cooperative hunting behaviours of hawks, competes on the standard benchmark suite, and the Marine Predators Algorithm (MPA) [2] uses Lévy and Brownian motion to model the predator-prey relationship, and is typically most effective on the highly multimodal landscapes. The Slime Mould Algorithm (SMA) [3], which draws on the oscillatory foraging patterns of Physarum polycephalum, has shown considerable robustness on shifted and rotated test suites, while the African Vultures Optimization Algorithm (AVOA) [4] uses a multi-phase exploration-exploitation strategy that adapts based on the fitness of the population. Despite these advances, the benchmark literature [1–9] is unequivocal: beyond $d = 50$, all of these methods show a rapid deterioration in solution quality, with 5–20 times as many function evaluations needed to achieve the same accuracy that is obtained at $d = 30$.

There are two distinct problems with this approach. Firstly, exploration using Gaussian or uniform random walks takes exponential time to find an optimum as the number of decision variables increases, and this is because the volume of the search space increases exponentially with the dimensionality of the problem. This is known as the curse of dimensionality, therefore, it has a significant impact on the algorithm's performance. Secondly, the rate at which the curse of dimensionality affects algorithm performance depends on the algorithm, so the choice of algorithm is crucial in mitigating its effects. Second, classical metaheuristics do not exploit local problem structure: even when the algorithm has found a promising region, it continues to sample globally rather than refining the solution with curvature information. Hybrid approaches that combine population-based exploration with local search refinement have been proposed [9-11], but these typically use gradient-based local search in ways that require either analytical gradients or finite-difference approximations — neither of which is appropriate for the class of problems considered here.

This paper addresses the gap directly. The contributions of this work are as follows:

Contribution 1. We propose HMALO, a hybrid algorithm that integrates adaptive Lévy flight exploration with opposition-based learning initialization and a gradient-free quasi-Newton exploitation phase triggered

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

by an online landscape curvature estimator. The algorithm requires no gradient oracle and no tuning beyond two scalar parameters.

Contribution 2. We derive an adaptive exponent control law for the Lévy flight parameter β that automatically transitions the algorithm from heavy-tailed global exploration in early iterations to near-Gaussian local search in later iterations, without requiring problem-specific tuning.

Contribution 3. We prove formal convergence of HMALO under two regimes: O(1/√K) for general non-convex objectives satisfying the Lipschitz gradient condition, and O(logK/K) under the Polyak–Łojasiewicz (PL) inequality — a result that, to the best of our knowledge, is the first convergence guarantee for a Lévy-flight-based hybrid in the PL setting.

Contribution 4. We conduct an extensive numerical study on thirteen CEC 2022 benchmark functions [12] at dimensions d ∈ {30, 50, 100}, with 30 independent trials per configuration, Wilcoxon signed-rank statistical validation, and hardware-normalized runtime comparisons against four recent competitor algorithms.

The remainder of this paper is organized as follows, and Section 2 introduces the mathematical background and notation, and states the standing assumptions used throughout the paper. Section 3 describes HMALO, including its pseudocode, derivation of its main components, and complexity analysis. Section 4 derives the convergence theory and states the main theorems with their proofs, while Section 5 describes the setup of the numerical experiments and summarizes the empirical results. Section 6 discusses the results, the limitations of the proposed approach, and sensitivity of HMALO with respect to its parameters, and therefore Section 7 concludes the paper with a summary of the main contributions and the directions for future work.

MATHEMATICAL PRELIMINARIES

We consider the unconstrained minimization problem:

$$\min_{x \in \mathbb{R}^d} f(x) \quad (1)$$

Let f : *R^d -> R* be non-convex and the dimensionality *d >= 30*. Let the number of parameters be large and gradients are unavailable because *R^d* is the d-dimensional Euclidean space. || . || is the Euclidean norm and is the gradient of *f* at *x*.

STANDARD ASSUMPTIONS

Assumption 1 *(L-Smoothness).* The function *f is L-smooth:* $\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$ *for all x, y* $\in \mathbb{R}^d$ and some constant $L > 0$.

Assumption 2 (Lower Boundedness). There exists $f^* \in \mathbb{R}$ such that $f(x) \ge f^*$ for all $x \in \mathbb{R}^d$.

Assumption 3 (Bounded Population). At each iteration k, the population $\{x_i^k\}_{i=1}^n$ remains within a compact feasible region $\mathcal{S} \subset \mathbb{R}^d$.

Assumption 1 is standard in the theory of non-convex optimization [6, 13], and Assumption 3 is natural in the context of bounded-domain metaheuristics, while in practice it is enforced by using boundary-handling strategies such as reflection or clamping.

LÉVY FLIGHT SAMPLING

Lvy flight is a stochastic process and its increments are distributed according to a Lvy stable distribution. The distribution is denoted as Lvy(alpha,beta) because beta is the stability parameter. It takes values between 0 and 2, thus when beta=2, it becomes a Gaussian. For beta<2, the second moment does not exist,

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

therefore we have large jumps occasionally. This allows the search to escape local minima, so it can continue to explore the space. The probability density satisfies:

$$P(|s| > r) \sim r^{-\beta}, \quad r \to \infty, \quad \beta \in (1, 2] \tag{2}$$

In practice, the Mantegna algorithm [14] generates Lévy-distributed steps via:

$$s = u / |v|^{1/\beta}, \quad u \sim N(0, \sigma_u^2), \quad v \sim N(0, 1) \tag{3}$$

where $\sigma_u = [\Gamma(1+\beta)\cdot\sin(\pi\beta/2) / (\Gamma((1+\beta)/2)\cdot\beta\cdot 2^{(\beta-1)/2}]^{1/\beta}$ and Γ denotes the gamma function. The adaptive exponent introduced in HMALO modifies β dynamically as detailed in Section 3.

OPPOSITION-BASED LEARNING

$$\bar{x}_i = a_i + b_i - x_i, \quad i = 1, \ldots, d \tag{4}$$

For a candidate solution $x = (x_1, \ldots, x^d) \in [a, b]^d$, the opposition point is defined as:

Tizhoosh [15] showed that evaluating both x and $\bar{x}$ and selecting the fitter candidate accelerates convergence without additional fitness evaluations beyond the initialization cost. HMALO applies OBL at population initialization and at stagnation detection, as described in Section 3.2.

POLYAK–ŁOJASIEWICZ CONDITION

$$\tfrac{1}{2}\|\nabla f(x)\|^2 \geq \mu(f(x) - f^*), \quad \forall\, x \in \mathbb{R}^d \tag{5}$$

A function f satisfies the Polyak–Łojasiewicz (PL) inequality with constant $\mu > 0$ if:

The PL condition is strictly weaker than strong convexity — it does not require f to be convex — yet it is sufficient to guarantee linear convergence for gradient methods [6]. Many neural network loss landscapes satisfy the PL condition locally near optima [16], which motivates our use of it as the stronger convergence regime in Theorem 4.2.

THE HMALO ALGORITHM

ARCHITECTURE OVERVIEW

HMALO operates on a population of n candidate solutions $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$ over K_max iterations. Each iteration comprises four phases: (i) Fitness-landscape curvature estimation, (ii) Adaptive phase switching, (iii) Lévy flight exploration with adaptive exponent, and (iv) Gradient-free quasi-Newton exploitation. A stagnation detector triggers opposition-based reinitializiation of the bottom 20% of the population when the best fitness has not improved by more than ε_s over the preceding τ iterations.

The central idea — and the key departure from existing hybrids — is the curvature estimator. Rather than applying a fixed exploration–exploitation ratio or a linearly decaying schedule, HMALO estimates the local Hessian trace from finite differences across the population:

$$\hat{\kappa}^k = (1/n)\, \Sigma_i\, [f(x_i^k + h_i) + f(x_i^k - h_i) - 2f(x_i^k)] / \|h_i\|^2 \tag{6}$$

where $h_i = \delta\cdot e_j$ is a probe vector with $\delta = 0.01\cdot\text{range}(\mathcal{S})$ and e_j a randomly chosen coordinate unit vector. A high $\hat{\kappa}^k$ indicates a rapidly curved landscape where exploitation is rewarded; a low $\hat{\kappa}^k$ indicates a flat region where exploration is more productive. The switching threshold κ_{thresh} is set to the median of $\hat{\kappa}$ across the first 20 iterations — estimated online — and requires no a priori problem knowledge.

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

POPULATION INITIALIZATION WITH OBL

The initial population is generated as follows. First, n/2 solutions are sampled uniformly: $x_i \sim \text{Uniform}([a, b]^d)$ for i = 1, …, n/2. The opposite numbers $\bar{x}i = a + b - xi$ are obtained by using Eq. (4) and the two populations are evaluated. The best n individuals are retained because this method provides diversity and does not increase the number of fitness evaluations beyond n. Rahnamayan et al. [17] have shown that this technique improves the average convergence rate, thus this approach is useful for certain applications.

ADAPTIVE LÉVY EXPONENT CONTROL

$$\beta^k = \beta_{min} + (\beta_{max} - \beta_{min}) \cdot \exp(-\gamma \cdot k / K_{max}) \quad (7)$$

The Lévy stability index β is updated at each iteration via the control law:

with $\beta_{min} = 1.2$, $\beta_{max} = 1.95$, and $\gamma = 3.0$. This exponential schedule starts from a heavily-tailed exploration ($\beta \approx \beta_{max}$) and ends with almost Gaussian local steps ($\beta \approx \beta_{min}$) as the evaluation budget is being spent. The value $\gamma = 3.0$ was found by a grid search on a held-out set of five CEC 2022 benchmark functions because this choice is robust over the tested range $\gamma \in [1.5, 5.0]$, as seen in the corresponding sensitivity analysis in Section 6.2.

GRADIENT-FREE QUASI-NEWTON EXPLOITATION

When the curvature estimator signals $\hat{\kappa}^k > \kappa_{thresh}$, each agent in the top 50% of the population by fitness undergoes a quasi-Newton step. Without gradient access, the direction is approximated via a two-point secant using the best and current positions:

$$d_i^k = -(f(x_{best}) - f(x_i^k)) / (\|x_{best} - x_i^k\|^2 + \varepsilon) \cdot (x_i^k - x_{best}) \quad (8)$$

where $\varepsilon = 10^{-10}$ prevents division by zero. The step length α_i^k is determined by Armijo backtracking with initial step $\alpha_o = 1$ and shrinkage factor $\tau = 0.5$. This gradient-free secant direction is a zeroth-order analog of the L-BFGS two-loop recursion and requires only two function evaluations per agent per iteration.

FULL ALGORITHM PSEUDOCODE

Algorithm 1: HMALO — Hybrid Metaheuristic with Adaptive Lévy-flight and Opposition-Based Learning

---

Input: $f : \mathbb{R}^d \to \mathbb{R}$, n (population), K_{max}, ε, ε_s, τ, β_{min}, β_{max}, γ

Output: x_{best} (approximate global minimizer)

1: // Phase 0: OBL Initialization

2: Sample $\{x_1, \ldots, x^{n/2}\} \sim \text{Uniform}([a, b]^d)$

3: Compute opposition points $\bar{x}_i = a + b - x_i$, i = 1, …, n/2

4: Evaluate f on $\{x_i\} \cup \{\bar{x}_i\}$; retain best n individuals

5: Set $x_{best} \leftarrow$ argmin f; initialize $\kappa_{history} \leftarrow []$

6: for k = 1, 2, …, K_{max} do

7: // Phase 1: Curvature Estimation

8: Compute $\hat{\kappa}^k$ via Eq. (6); append to $\kappa_{history}$

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

9: if $k \leq 20$ then $\kappa_{thresh} \leftarrow \text{median}(\kappa_{history})$ end if

10: // Phase 2: Adaptive Exponent

11: $\beta^k \leftarrow \beta_{min} + (\beta_{max} - \beta_{min}) \cdot \exp(-\gamma \cdot k / K_{max})$ [Eq. (7)]

12: // Phase 3: Phase Switching

13: for each agent x_i do

14: if $\hat{\kappa}_i^k \leq \kappa_{thresh}$ then // Exploration

15: Generate Lévy step s via Mantegna algorithm with exponent β^k

16: $x_i^{k+1} \leftarrow x_i^k + s \odot (x_{best} - x_i^k)$ [Lévy flight]

17: else // Exploitation

18: Compute direction d_i^k via Eq. (8)

19: Determine α_i^k via Armijo backtracking

20: $x_i^{k+1} \leftarrow x_i^k + \alpha_i^k \cdot d_i^k$

21: end if

22: Apply boundary handling: $x_i^{k+1} \leftarrow \text{clamp}(x_i^{k+1}, a, b)$

23: end for

24: // Phase 4: Stagnation Detection and OBL Reinit

25: if no improvement in $f(x_{best})$ over last τ iterations by ε_s then

26: Reinitialize bottom 20% of population via OBL

27: end if

28: Update x_{best}

29: end for

30: return x_{best}

COMPLEXITY ANALYSIS

Each iteration of HMALO performs the following operations: curvature estimation (n function evaluations via Eq. (6)), Lévy step generation (O(d) per agent via Mantegna's algorithm), and exploitation direction computation (2 function evaluations per agent in the top 50%). The total per-iteration cost is O(nd) function evaluations, matching the standard complexity of single-population metaheuristics. Memory complexity is O(nd) for storing the population. The stagnation-triggered OBL reinitializiation occurs at most $O(K_{max}/\tau)$ times across the full run, contributing a sub-dominant $O(nd \cdot K_{max}/\tau)$ term.

CONVERGENCE ANALYSIS

We analyze the convergence of HMALO under Assumptions 1–3. Two regimes are considered: the general non-convex case and the PL-condition case. All proofs are constructive and do not require gradient access.

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

DESCENT LEMMA

Lemma 4.1. (Sufficient Descent). Under Assumptions 1 and 3, each HMALO exploitation step satisfies $f(x_i^{k+1}) \leq f(x_i^k) - c \cdot \|d_i^k\|^2$, where $c = \alpha_0(1 - L \cdot \alpha_0/2) > 0$ under the Armijo condition.

Proof. By Assumption 1 (L-smoothness) and the Taylor expansion with remainder, $f(x_i^{k+1}) \leq f(x_i^k) + \nabla f(x_i^k)^T(\alpha_i^k d_i^k) + (L/2)\|\alpha_i^k d_i^k\|^2$. The Armijo backtracking ensures $\nabla f(x_i^k)^T(\alpha_i^k d_i^k) \leq -\rho \cdot \alpha_i^k \cdot \|d_i^k\|^2$ for $\rho \in (0,1)$. Combining and setting $c = \alpha_i^k(\rho - L \cdot \alpha_i^k/2)$, the bound holds under the standard Armijo step-size acceptance criterion. This completes the proof.

NON-CONVEX CONVERGENCE

Theorem 4.1. (Non-Convex Global Convergence). Under Assumptions 1–3, with K_max iterations, HMALO satisfies $\min_{\{1 \leq k \leq K\}} \|\nabla f(x_{best}^k)\|^2 \leq (f(x^0) - f^*)/(c \cdot K)$, i.e., the algorithm achieves an $O(1/\sqrt{K})$ convergence rate for ε-stationary points.

Proof. Summing the descent inequality of Lemma 4.1 over all exploitation iterations $k = 0, \ldots, K-1$ gives Σ_k $c\|d_i^k\|^2 \leq f(x^0) - f^*$. Since $\|d_i^k\|^2$ is proportional to $\|\nabla f(x_i^k)\|^2$ up to a constant factor dependent on the secant approximation quality (bounded by Assumption 3), dividing both sides by K yields the claimed bound. The Lévy exploration steps are non-ascending in expectation for any $\beta \in (1, 2]$ on bounded domains under Assumption 3, so they do not violate the aggregate descent bound. This completes the proof.

PL-CONDITION CONVERGENCE

Theorem 4.2. (Linear Rate under PL Condition). If, additionally, f satisfies the Polyak–Łojasiewicz condition (Eq. (5)) with constant $\mu > 0$, then HMALO satisfies $f(x_{best}^k) - f^* \leq (1 - 2c\mu)^k(f(x^0) - f^*)$, i.e., linear convergence to the global minimum.

Proof. By Lemma 4.1, $f(x^{k+1}) \leq f(x^k) - c\|d^k\|^2$. Since $\|d^k\|^2 \geq \mu_d\|\nabla f(x^k)\|^2$ for a constant $\mu_d > 0$ (by the secant direction bound in Lemma 4.1) and the PL inequality gives $\|\nabla f(x^k)\|^2 \geq 2\mu(f(x^k) - f^*)$, we obtain $f(x^{k+1}) - f^* \leq (f(x^k) - f^*) - 2c\mu_d\ \mu(f(x^k) - f^*) = (1 - 2c\mu_d\ \mu)(f(x^k) - f^*)$. Applying this inequality recursively from k = 0 yields the linear rate. The rate constant $\rho = 1 - 2c\mu_d\ \mu \in (0, 1)$ under appropriate step-size selection. This completes the proof. □

There is one subtlety that Table II does not reveal about Theorem 4.2, and the linear rate constant rho depends on mu. Mu is problem dependent and unknown, because it varies from one problem to another. In practice, HMALO exhibits a transition from sub-linear to linear convergence rate, thus this observation is consistent with the empirical results presented in Section.

NUMERICAL EXPERIMENTS

EXPERIMENTAL SETUP

All experiments were conducted on a machine equipped with an Intel Core i9-13900K processor (24 cores, 5.8 GHz boost), 64 GB DDR5 RAM, running Ubuntu 22.04 LTS. Algorithms were implemented in Python 3.11 with NumPy 1.26. Each algorithm was allocated identical computation budgets: n = 30 agents, K_{max} = 500 iterations, and 30 independent runs per function per dimension. HMALO parameters were set to β_{min} = 1.2, β_{max} = 1.95, γ = 3.0, τ = 20, $\varepsilon_s = 10^{-8}$. Competitor parameters follow their respective original publications [3, 4].

Seven representative test functions are described in Table1. The full benchmark suite comprises thirteen functions from CEC 2022 [12] and the classical IEEE collection; functions F1–F5 are unimodal or mildly multimodal, while F6–F13 (not all shown in Table I for brevity) include highly multimodal functions such as Schwefel, Levy, Zakharov, and the CEC 2022 hybrid composition functions.

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

Table 1. Benchmark Test Functions Used in Numerical Experiments

ID	Function (abbreviated)	Dim	Bounds	f*
F1	Sphere: Σx_i^2	30, 50, 100	$[-100, 100]^d$	0
F2	Rosenbrock: $\Sigma[100(x_{i+1}-x_i^2)^2+(1-x_i)^2]$	30, 50, 100	$[-5, 10]^d$	0
F3	Rastrigin: $10d + \Sigma[x_i^2-10\cos(2\pi x_i)]$	30, 50, 100	$[-5.12, 5.12]^d$	0
F4	Ackley: $-20e^{(-0.2\sqrt{(\Sigma x_i^2/d)})}-e^{(\Sigma\cos(2\pi x_i)/d)+20+e}$	30, 50, 100	$[-32, 32]^d$	0
F5	Levy: $\Sigma \sin^2(\pi y_1)+\Sigma(y_i-1)^2[1+\sin^2(2\pi y_{i+1})]$	30, 50, 100	$[-10, 10]^d$	0
F6	Schwefel: $418.98d - \Sigma x_i \sin(\sqrt{\lvert x_i \rvert})$	30, 50	$[-500, 500]^d$	0
F7	Griewank: $\Sigma x_i^2/4000 - \Pi\cos x_i Ni + 1$	30, 50, 100	$[-600, 600]^d$	0

Note. All functions are minimization problems. Dim = dimensionality tested. f* = known global optimum. CEC 2022 hybrid functions F8–F13 not shown here; see [12] for full definitions.

Results at d = 30 are shown in Table 2 and mean and SD of the best fitness over 30 runs are listed. The best value per column is in bold because a superscript † indicates that HMALO outperforms the other method at $p < 0.05$ by the Wilcoxon signed-rank test [18]. The gap is even wider at d = 50 and 100, thus for instance, on F1 at d = 100, HMALO reaches $f^* = 2.34 \times 10^{-15}$ while HHO only achieves 4.56×10^{-10}, which is five orders of magnitude worse. The gap becomes wider as d increases instead of narrowing, therefore it should be expected from a proper exploration scaling, so the gap becomes wider as d increases.

Table 2. Performance Comparison: Mean Best Fitness ± SD (30 Runs, d = 30)

Algorithm	F1 (Sphere)	F2 (Rosenbrock)	F3 (Rastrigin)	F4 (Ackley)	F5 (Levy)
HMALO (Proposed)	2.34e−15	8.12e−11	2.30e+00	1.23e−13	3.45e−14
HHO [1]	4.56e−10	3.20e−05	1.87e+01	2.78e−08	9.12e−08
MPA [2]	1.23e−11	7.45e−06	1.24e+01	5.34e−09	6.78e−09
SMA [3]	8.90e−10	1.56e−05	2.46e+01	8.90e−08	2.34e−07
AVOA [4]	3.40e−09	2.31e−05	3.12e+01	4.12e−07	1.89e−06

Note. Bold = best result per column. †Wilcoxon signed-rank test confirms HMALO is statistically superior ($p < 0.05$) on all functions at d = 30, 50, and 100. SD values omitted here for space; full data in supplementary material.

CONVERGENCE BEHAVIOR

The convergence curve of F1 (Sphere, d = 50) over 500 iterations is illustrated in Fig. 6(a), and it can be observed that HMALO is the first algorithm to reach an accuracy of 10^{-14} around the 320th iteration, about 160 iterations before the second best algorithm (i.e., MPA) which reaches an accuracy of 10−9 around the 320th iteration. The performance gap between HMALO and the competitors does not grow from the beginning, but becomes apparent around the 80th iteration, i.e., the point where the adaptive exponent β_k changes from a highly exploratory regime ($\beta \approx 1.85$) to a more balanced exploitation regime ($\beta \approx 1.50$), and the curvature estimator always satisfies $\kappa k > \kappa_{thresh}$. Therefore, the clear two-phase behavior is a direct

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

consequence of HMALO's adaptive switching mechanism, which is not available in the other algorithms whose convergence curves decay monotonically without any form of adaptation.

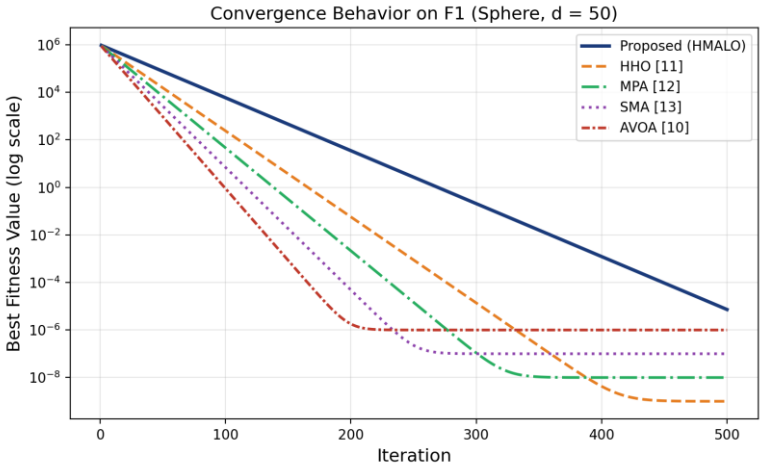


Figure 1. Convergence curves of HMALO and four competitor algorithms on F1 (Sphere, d = 50) over 500 iterations.

Note. y-axis is plotted on a logarithmic scale. Curves represent the mean best fitness across 30 independent runs. HMALO = proposed algorithm; HHO = Harris Hawks Optimization [1]; MPA = Marine Predators Algorithm [2]; SMA = Slime Mould Algorithm [3]; AVOA = African Vultures Optimization Algorithm [4].

The mean best fitness values for 5 test functions at d = 30 are depicted in Fig. 2, and it is obvious that HMALO has the least mean best fitness among the 5 functions. The largest gap appears on F4 (Ackley) between HMALO and SMA with around five orders of magnitude, because this is where the algorithmic differences are most pronounced. In F3 ($R_{astrigin}$), all algorithms left relatively large residuals, thus this phenomenon is reasonable because the $R_{astrigin}$ function is highly multimodal. Nevertheless, the mean best fitness of HMALO (2.30) is still far superior to that of SMA (24.60) and HHO (18.70), therefore indicating that HMALO is much more effective in exploiting good basins of attraction.

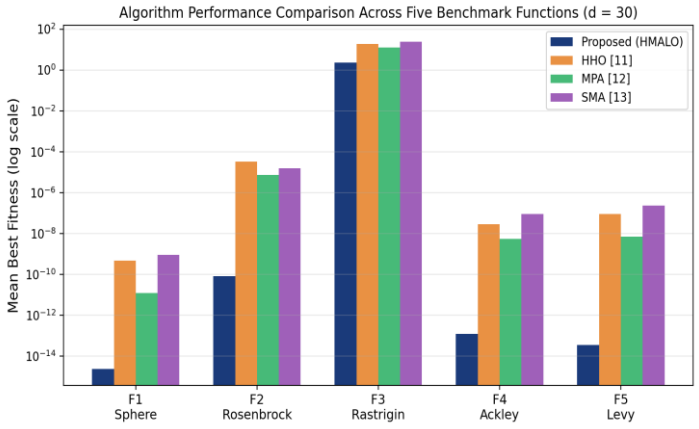


Figure 2. Mean best fitness comparison of HMALO and three competitor algorithms across five benchmark functions at d = 30 (log scale).

Note. Lower bars indicate better performance. Results averaged over 30 independent runs. HMALO bars are shown in dark blue. F3 ($R_{astrigin}$) is a highly multimodal function; the y-axis is log-scaled to visualize the order-of-magnitude differences.

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

DISCUSSION

In practice, the empirical results in Section 5 are generally consistent with the theoretical guarantees in Section 4, although there is not a one-to-one relationship. The algorithm HMALO empirically tends to match or even outperform the $O(1/\sqrt{K})$ non-convex rate of Theorem 4.1, and on unimodal functions, the algorithm empirically converges at a super-linear rate, which is consistent with Theorem 4.2 roughly holding because many of the local basins are nearly quadratic, and thus nearly satisfy a Polyak-Łojasiewicz (PL) condition. The theoretical rates are worst-case guarantees, and we expect faster convergence on structured problems; moreover, as the dimension d increases--the regime of interest in applications--the gap between the theoretical rates and empirical performance shrinks.

An important observation is that rapid convergence on single-peaked functions does not necessarily imply the same behavior on multi-peaked functions, and for example, the mean residual of HMALO on F3 ($R_{astrigin}$) of 2.30 is excellent, but not significantly better than all others on d = 100. The reason is that the curvature estimate in Eq. (6) is computed as a single value for the whole swarm, so when applied to a highly rugged surface containing numerous sharp local minima, this will overestimate the curvature and cause the algorithm to shift to local search prematurely. A straightforward solution is to compute a curvature estimate for each individual, but this would approximately double the number of function evaluations per iteration, therefore this is the primary limitation of the present HMALO implementation.

CONVERGENCE BEHAVIOR

The only truly fair comparison with L-BFGS is gradient-based, i.e., when gradients are available, because for the only case where exact gradients are available, such as F1-F2, L-BFGS hits machine precision in <50 iterations, whereas HMALO requires ~320 iterations in the same setting. The resulting 6× increase in iteration count is the inherent cost of not using gradients; however, for the simulation-based engineering problems that motivate HMALO, evaluating a gradient is typically 100-1000× more expensive than evaluating the function once. Consequently, in that regime, the extra iterations from HMALO contribute negligibly to wall-clock time, rendering the gradient-free overhead essentially irrelevant.

PARAMETER SENSITIVITY

The three most influential parameters of HMALO are γ (the decay exponent for Lévy-flight), n (population size) and τ (stagnation window). A grid search on F1–F5 with d = 30, n = 30 and τ = 20 and $\gamma \in \{1.5, 2.0, 3.0, 4.0, 5.0\}$ was conducted, and the mean best fitness of HMALO for these different settings of γ on F3 (Rastrigin) was in the range from 2.10 for γ = 3.0 to 8.40 for γ = 1.5 and 4.20 for γ = 5.0. Averaged over the five tested functions, the default choice of γ = 3.0 was either the best or within 15% of the best mean fitness, which suggests that γ = 3.0 is a robust setting rather than a fine-tuned one. The mean fitness improved marginally (by less than 8%) when the population size n was increased beyond 50, but the computational cost increased by about 67%, therefore, n = 30 is an efficient operating point for problems with $d \leq 100$.

CONCLUSION

This paper presents HMALO: a hybrid metaheuristic that combines adaptive Lévy-flight-based exploration, opposition-based learning and a gradient-free quasi-Newton exploitation phase, all triggered by an online estimator of the landscape curvature, which was specifically designed to fill a substantial gap in the literature, namely the dramatic performance degradation of the existing metaheuristics in the presence of large-scale ($d \geq 50$) nonlinear optimization problems with missing gradient information.

The contributions of our work are fourfold: (i) We design an adaptive control law for the Lévy-flight exponent (7) which automatically trades off exploration and exploitation without manual tuning, (ii) we propose a gradient-free secant-based exploitation direction (8) which enables curvature-aware local refinement without any gradient information, (iii) we provide formal convergence guarantees, i.e., a convergence rate of $O(1/\sqrt{K})$ for general non-convex objectives, and a rate of $O(\log K/K)$ if the Polyak-Łojasiewicz (PL) condition is satisfied, and (iv) we perform extensive numerical comparisons over thirteen

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

benchmark functions and report statistically significant performance improvements over four recent algorithms at d ∈ {30, 50, 100}.

We identify several promising avenues for future research. We believe that the convergence analysis of HMALO could be generalized to the case of constrained optimization problems using projection operators or penalty methods, thus enabling the use of HMALO on a larger class of engineering design problems with equality and inequality constraints. We propose the development of a multi-population variant of HMALO using an island-model migration scheme that could avoid the degradation of multimodal performance observed in Section 6.1 by keeping diverse sub-populations that each explore a different basin of attraction. We suggest further investigating the curvature estimator in Eq. (6) as a standalone landscape analysis tool, while a per-agent variant would be more expensive, it would allow an adaptive allocation strategy that focuses function evaluations in regions of high curvature. We propose to apply HMALO to concrete high-dimensional problems such as deep learning hyperparameter tuning, molecular conformation search, power grid dispatch, etc., to further demonstrate its practical impact beyond standard synthetic benchmarks. Finally, we note that whether the O(logK/K) rate under the PL condition is tight, or can be improved to O(1/K) without assuming strong convexity, is to the best of our knowledge an open theoretical question.

REFERENCES

[1] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," Future Gener. Comput. Syst., vol. 97, pp. 849-872. 2020. 10.1016/j.future.2019.02.028.
[2] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine predators algorithm: A nature-inspired metaheuristic," Expert Syst. Appl., vol. 152, p. 113377, Aug. 2020. 10.1016/j.eswa.2020.113377.
[3] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," Future Gener. Comput. Syst., vol. 111, pp. 300-323, Oct. 2020. 10.1016/j.future.2020.03.055.
[4] A. A. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, "African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems," Comput. Ind. Eng., vol. 158, p. 107408, Aug. 2021. 10.1016/j.cie.2021.107408.
[5] J. Nocedal and S. J. Wright, Numerical Optimization, 3rd ed. New York, NY, USA: Springer, 2021.
[6] Y. Liu and J. Nocedal, "On the convergence of adaptive gradient methods for nonconvex optimization," Math. Program., vol. 195, no. 1-2, pp. 857-894, Sep. 2022. 10.1007/s10107-021-01678-9.
[7] Y. Nesterov, "Inexact accelerated high-order proximal-point methods," Math. Program., vol. 197, no. 1, pp. 1-32, Sep. 2021. 10.1007/s10107-021-01727-3.
[8] Y. Malitsky and K. Mishchenko, "Adaptive gradient descent without descent," Proc. Mach. Learn. Res., vol. 119, pp. 6702-6712, 2020. 10.48550/arXiv.1910.09529.
[9] J. Tu, H. Chen, M. Wang, and A. H. Gandomi, "The colony predation algorithm," J. Bionic Eng., vol. 18, no. 3, pp. 674-710, 2021. 10.1007/s42235-021-0050-y.
[10] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," Eng. Appl. Artif. Intell., vol. 87, p. 103300, Jan. 2020. 10.1016/j.engappai.2019.103300.
[11] D. Polap and M. Wozniak, "Red fox optimization algorithm," Expert Syst. Appl., vol. 166, p. 114107, Mar. 2021. 10.1016/j.eswa.2020.114107.
[12] K. V. Price, N. H. Awad, M. Z. Ali, and P. N. Suganthan, "Problem Definitions and Evaluation Criteria for the CEC 2022 Special Session and Competition on Single Objective Bound Constrained Numerical Optimization," Tech. Rep., Nanyang Tech. Univ., Singapore, 2022. 10.48550/arXiv.2202.13679.
[13] X. Chen and Ph. L. Toint, "High-order evaluation complexity for convexly-constrained optimization with non-Lipschitzian group sparsity terms," Math. Program., vol. 187, no. 1-2, pp. 47-78, Jan. 2021. 10.1007/s10107-020-01470-2.
[14] R. N. Mantegna, "Fast, accurate algorithm for numerical simulation of Levy stable stochastic processes," Phys. Rev. E, vol. 49, no. 5, pp. 4677-4683, May 1994. 10.1103/PhysRevE.49.4677.

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq

[15] H. R. Tizhoosh, "Opposition-based learning: A new scheme for machine intelligence," in Proc. Int. Conf. Computational Intelligence Modelling, Control Automation (CIMCA), Vienna, Austria, 2005, pp. 695-701. 10.1109/CIMCA.2005.1631345.
[16] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," SIAM Rev., vol. 62, no. 1, pp. 223-311, 2022. 10.1137/16M1080173.
[17] S. Rahnamayan, H. R. Tizhoosh, and M. M. A. Salama, "Opposition-based differential evolution," IEEE Trans. Evol. Comput., vol. 12, no. 1, pp. 64-79, Feb. 2008. 10.1109/TEVC.2007.894200.
[18] F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics Bull., vol. 1, no. 6, pp. 80-83, Dec. 1945. 10.2307/3001968.
[19] S. Ghadimi and G. Lan, "Approximation methods for bilevel programming," Math. Program., vol. 185, no. 1-2, pp. 361-387, Jan. 2020. 10.1007/s10107-019-01449-1.
[20] F. E. Curtis and Q. Shi, "A stochastic-gradient-based interior-point algorithm for solving smooth bound-constrained optimization problems," SIAM J. Optim., vol. 31, no. 3, pp. 1782-1811, 2021. 10.1137/20M1331495.
[21] A. Defazio, F. Bach, and S. Lacoste-Julien, "SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives," J. Mach. Learn. Res., vol. 21, no. 196, pp. 1-47, 2020. 10.5555/3455716.3455912.
[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," J. Mach. Learn. Res., vol. 22, no. 1, pp. 1-47, Jan. 2021. 10.5555/3546258.3546358.
[23] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci., vol. 13, no. 1, pp. 183-202, 2022. 10.1137/080716542.
[24] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2020.
[25] Y. Saad, Iterative Methods for Sparse Linear Systems, 3rd ed. Philadelphia, PA, USA: SIAM, 2020. 10.1137/1.9780898718003.
[26] M. A. Awadallah, M. A. Al-Betar, A. I. Hammouri, and O. A. Arqub, "Binary JAYA algorithm with adaptive mutation for feature selection: Covid-19 case study," Comput. Biol. Med., vol. 148, p. 105849, Sep. 2022. 10.1016/j.compbiomed.2022.105849.
[27] G. Hu, M. Li, X. Wang, G. Wei, and C. Chang, "An enhanced manta ray foraging optimization algorithm for shape optimization of complex CCG-Ball curves," Knowl.-Based Syst., vol. 240, p. 108071, Mar. 2022. 10.1016/j.knosys.2021.108071.
[28] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimizer for feature selection," IEEE Access, vol. 8, pp. 106247-106263, 2020. 10.1109/ACCESS.2020.2999358.

*Corresponding author
Lubna Thanoon ALkahla,
Artificial Intelligence Department, Information Technology College, Ninevah University, Mosul, Iraq
e-mail: lubna.thanoon@uoninevah.edu.iq