UNDER PEER REVIEW

On the Rate Convergence of two Particle Swarm Optimization Algorithms: Gradient-Perturbation and Dual-Binary

Abstract

Particle Swarm Optimization (PSO) is a widely used metaheuristic for solving complex optimization problems, yet its theoretical convergence properties remain an active area of research. This paper introduces a novel criterion for evaluating the rate of convergence of PSO algorithms, providing a new perspective on their theoretical efficiency. Using this criterion, we establish rigorous mathematical results that extend existing analyses of PSO convergence.

A key contribution of this work is the introduction of a stochastic dynamic averaging technique to bound the approximation error, offering deeper insights into the behavior of PSO algorithms. Specifically, we analyze two variants: Gradient-Perturbation PSO (GP-PSO) and Dual-Binary PSO (DB-PSO). By leveraging functional analysis and probability theory, we derive improved convergence guarantees and demonstrate that these methods provide more precise accuracy bounds compared to classical approaches. The combination of rigorous theoretical analysis and empirical validation strengthens the understanding of PSO's efficiency and provides new insights into its convergence behavior.

Keywords: Approximation, stochastic modelling, gradient perturbation, optimization

1. Introduction and problem setting

We propose a novel standard for assessing the theoretical efficacy of swarm algorithms. Usually, the standard relies on the worst error of the algorithm , which does not account for the rate of convergence. Recently this worst error has been analyzed by some authors

[12,13,14]. They have highlighted the case of a multivariate approximation problems for functions of n variables from the Hilbert space. The squared-exponential reproducing kernel (SERK) for every $n \in \mathbb{N}$ as $n \to \infty$ is given by

$$\mathcal{K}_n(t_1, ..., t_n; s_1, ..., s_n) = \prod_{m=1}^{n} \exp\left\{-\gamma^2(t_m - s_m)^2\right\}, \quad (1)$$

where $t = (t_1, ..., t_n)$ and $s = (s_1, ..., s_n)$ are from $\mathbb{R}^n$ ($\mathbb{R}$ is the set of real numbers), $\gamma > 0$ is a shape parameter. The Hilbert space $H_{n,\gamma}$ with the above SERK is well studied and it is used widely in numerical computations, statistical learning and engineering. We consider $L_{2,n}$, the space of functions that have the finite norm

$$||f|| := \sum_{m=1}^{n} ||f||_{L_{2,m}} \quad (2)$$

where

$$||f_m||^2_{L_{2,n}} := \pi^{-\frac{1}{2}} \int_{\mathbb{R}^n} f^2(x) \prod_{m=1}^{n} e^{-x_m^2} dx \quad (3)$$

We consider the swarm multivariate approximation problem $SMAP_n : H_{n,\gamma} \to L_{2,n}$, and set $x = (x_1, ..., x_n)$ in the integral.

The accuracy of swarm algorithms is commonly evaluated using stochastic approximation methods. Previous research has proposed the use of stochastic approximation (SA) with PSO to improve performance or parameter selection [9,10].Our results demonstrate that some stochastic development methods are optimal in proving the lower bounds for $\phi_i(\tau, H)$ in the case where $\tau \leq \frac{(2-\beta)t}{(2-\beta)t+2}$. We use Hilbert spaces to describe systems where inner products and distances are naturally defined. The standard Particle Swarm Optimisation (PSO) algorithm, has been proposed by Kennedy and Eberhart in 1995 [4]. Let N be the dimension of the search space, and M be the individual size of the particle group. The current position of the $i-th$ particle is represented by $X_i = (x_{i_1}, x_{i_2}, ..., x_{i_N})$, and the current velocity is $V_i = (v_{i_1}, v_{i_2}, ..., v_{i_N})$. The current position of the $i-th$ particle is represented by $X_i = (x_{i_1}, x_{i_2}, ..., x_{i_N})$, and the current velocity is $V_i = (v_{i_1}, v_{i_2}, ..., v_{i_N})$. The particle's current position is $P_i = (p_{i_1}, p_{i_2}, ..., p_{i_N})$. For the entire particle swarm, a global optimal solution of $G(t) = (g_{t_1}, g_{t2}, ..., g_{tN})$ is obtained. The velocity and position update formulas for each iteration are given below:

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (4)$$

$$V_i(t+1) = \omega V_i(t) + c_1 r_1 (P_i(t) - X_i(t)) + c_2 r_2 (G(t) - X_i(t)) \quad (5)$$

$t = 0, 1, 2, ...; i = 1, 2, ..., M$. Here ω represents the inertia weight, balancing the algorithm's global search and local search ability. c_1 and c_2 denote individual cognitive social factors, respectively. r_1 and r_2 are random variables ranging from 0 to 1.

The Particle Swarm Optimization (PSO) approach exhibits slow convergence speed, low optimization accuracy and premature convergence when applied to complex functions, despite its advantages of simplicity, few parameters and ease of implementation. Instead of searching the entire parameter space, the particles are usually restricted to exploration around global and local optimums. Given the limitations of the standard

PSO algorithm, several authors have proposed numerous extensions [1, 2, 3]. To guarantee the stability and generate higher quality solutions than the basic PSO approach, the velocity is updated to $\chi \cdot V_{t+1}$, where $\chi = 2\theta^{-1}$, is the constriction factor and $\theta = |2 - \phi - \sqrt{\phi^2 - 4\phi}|; \phi = c_1 + c_2 > 4$. To evaluate the convergence rate, we focus on the gradient perturbation (GP-PSO) extension postulated by [6]. The GP-PSO formulas are presented below:

$$X_i(t+1) = X_i(t) + V_i(t+1) + \alpha_i(-\nabla_{X_i} f) \tag{6}$$

$$\phi_i = \frac{f(X_i) - f(X_i + \alpha_i d_i)}{f(X_i) - \Phi(X_i + \alpha_i d_i)} \tag{7}$$

where α_i in (4) can be calculated using the Wolfes rule, $\nabla_{X_i} f = \frac{\partial f}{\partial X_i}$ the Laplacian of f in X_i.

$d_i = -g_i(g_i = \nabla_{X_i})f)$; $\Phi(X_i + \alpha_i d_i) = f(X_i) + g_i^T(\alpha_i d_i)$; ϕ_i signifies the likeness amid the function $f(X_i + \alpha_i d_i)$ and $\Phi((X_i + \alpha_i d_i)$.

Here, $||g_i|| = \left(\alpha_i^{-1}[f(X_i) - \phi(X_i + \alpha_i d_i)]\right)^{\frac{1}{2}}$ and when $\alpha_i \to 0, \phi_i \to 1$. The algorithm's particular steps are outlined in Section 3 of reference [16]. We analyze swarm approximation with respect to a given dictionary (see definition below), and prove non-trivial inequalities for ϕ_i in both cases where E is a Hilbert space and a Banach space.

Let H denote a real Hilbert space with the inner product $< .,. >$ and norm $|| \cdot ||$. A set of elements (functions) $\mathcal{D}$ from H is considered a dictionary (symmetric dictionary) if each $g \in \mathcal{D}$ has a norm of one ($||g|| = 1$) and $\overline{span}\mathcal{D} = H$. For convenience, we additionally assume that $g \in \mathcal{D}$ implies $-g \in \mathcal{D}$, a property of symmetry.

To analyze the binary framework of PSO, the particle position is updated by toggling each bit value between 0 and 1 according to the velocity of that bit [18-III,16 paragraph 3.2]. To be more specific, for the $d-th$ bit of the $i-th$ particle, the velocity v_{id} is transformed (using the sigmoid function) into a probability, thus

$$P(V_i(t) = v_{id}) = \frac{1}{1 + e^{-v_{id}}}, \tag{8}$$

x_{id} takes 1 with a probability of $P(V_i(t) = v_{id})$. In this paper, velocity v_{id} is bounded by a threshold $\tilde{v}$ after being updated by equation (2). Thus,

$$v_{id} = \max(\tilde{v}, -\tilde{v})$$

By eliminating the bit index from (2):

$$V_{t+1} = \omega V_t + c_1 r_1(P - t - X_t) + c_2 r_2(G_t - X_t).$$

From there, it is evident that

$$P(X_t = 1) = \frac{1}{1 + e^{-V_t}} = 1 - P(X_t = 0) \tag{9}$$

If $0 < \omega < 1$, the function $E[V_{t+1} - V_t]$ decreases as V_t increases.

The search for the rate that minimizes ϕ_i in (4) is a fundamental theoretical problem in swarm approximation in Hilbert spaces [16, Paragraph 3.1]. It is evident that for any $X_t \in H$ such that $||X_t|| < \infty$,

$$||X_i(t+1) - X_i(t)|| \leq ||V_i(t+1) + \alpha_i(-\nabla_{X_i})F||$$

We aim to extend the asymptotic characteristics $\phi_i(H_t)$ for $\tau \in (0,1]$, define as follow:

$$\phi_i(\tau, H_t) := \inf \frac{||f(X_i) - f(X_i + \alpha_i d_i)||_{H_t}}{||(f(X_i)^{1-\tau} - \Phi(X_i + \alpha_i d_i)||^{\tau}_{H_t}} \tag{10}$$

Clearly

$$\phi_i(1, H_t) = \inf \frac{||f(X_i) - f(X_i + \alpha_i d_i)||_{H_t}}{||1 - \Phi(X_i + \alpha_i d_i||_{H_t}}$$

and $\phi_i(\tau, H) \geq \phi_i(\beta, H)$ if $\tau \leq \beta$. A comparison of 22 functions, in [16,table 1-2-3], provides information on the formation of modal functions and the performance of the GB-PSO algorithm. However, although this algorithm has a higher speed of convergence and stronger optimization capabilities, its convergence rate remains unclear. Therefore, we set up the boundaries as

$$\frac{1}{2} m^{-\frac{\tau}{2}} \leq \phi_m(\tau, H_t) \leq m^{-\frac{\tau}{2}}, \tau \leq \frac{1}{3}. \tag{11}$$

2. Main results

In this section we formulate the main results of the paper. The proofs are provided in section 4, the necessary auxiliary tools are presented in section 3.

We consider the convergence rate defined in the previous section. Let a parameter $\beta \in (0,1]$ and a sequence $\mu = \{u_m\}_{m=1}^{\infty}; 0 \leq u_m \leq 1$. We define the gradient swarm algorithm with parameter β.

We define $f_0 := f_0^{\mu,\beta} := f$. For each $m \geq 1$, we inductively define

- $\varphi_m := \varphi_m^{\mu,\beta} \in \mathcal{D}$ as any φ satisfying
$$< f_{m-1}, \varphi_m > \leq u_m \inf_{g \in \mathcal{D}} < f_{m-1}, g >$$
- $f_m := f_m^{\mu,\beta} := f_{m-1} - [\beta(2-\mu)]^m < f_{m-1}, \varphi_m > \varphi_m$
- $$S_m(f, \mathcal{D}) := S_m^{\mu,\beta}(f, \mathcal{D}) = \beta \sum_{j=1}^{m} < f_{j-1}, \varphi_j > \tag{12}$$

Now, we provide the necessary bound for $\phi_m^{(\mu,\beta)}(\tau, H_{n,\gamma})$ as

$$\phi_m^{(\mu,\beta)}(\tau, H_{n,\gamma}) = \inf_{\mathcal{D}} \inf_{f \in S_1(\mathcal{D}), f \neq 0} \inf_{S_m^{\mu,\beta}(f,\mathcal{D})} \frac{||f - S_m^{\mu,\beta}(f, \mathcal{D})||}{||f||^{1-\tau} ||f||^{\tau}_{S_1(\mathcal{D})}} \tag{13}$$

where $||f||_{S_1(\mathcal{D})} := \inf\{M > 0 : f/M \in S_1(\mathcal{D})\}$ for each $f \in H_{n,\gamma}$, and $S_1(\mathcal{D})$ is a natural occurring swarm class defined as a stochastic clustered group formed by closure of the nonconvex hull of $\mathcal{D}$.

Theorem 1. *In any Hilbert space $H_{n,\gamma}$,*

$$\phi_m^{\mu,\beta}(\tau, H_{n,\gamma}) \leq (1 + m\beta(2-\beta)\mu^2)^{-\frac{\tau}{2}}. \tag{14}$$

where τ_n is a sequence such that,

$$\tau_n \to \left(1 - \frac{\varphi_m(X_n + \alpha_n d_n)}{(\alpha_n ||g_n||)^2_m}\right)^{\frac{1}{2}}, n \to \infty \tag{15}$$

3. Auxiliary results

Lemma 1. *Let H be a Hilbert space, and $S_m^{\mu,\beta}$ be a swarm-based approximation operator. For any function $f \in H$, the following non-expansiveness properties holds:*

1. $||S_m^{\mu,\beta}(f,\mathcal{D})|| \leq ||f||$
2. $||f - S_m^{\mu,\beta}(f,\mathcal{D})|| \leq u_m||f||(1 + m\beta(2-\beta)\mu^2)^{-\frac{\tau}{2}}$

Proof of Lemma 1. By construction, the operator $S_m^{\mu,\beta}$ is defined as a weighted stochastic average of particles in the swarm. Let $\{X_k\}_{k=1}^m$ represent the position of the particles with a dynamic movement defined by

$$X_{k+1} = X_k - \mu\nabla f(X_k) + \beta(X_k - X_{best})$$

where X_{best} represents the best historical position. Taking norms on both sides and applying the triangle inequality,

$$||X_{k+1}|| \leq ||X_k|| + \mu\nabla f(X_k) + \beta||X_k - X_{best}||$$

Since X_{best} is chosen from the swarm $||X_k - X_{best}|| \leq ||X_k||$, and then

$$||S_m^{\mu,\beta}(f,\mathcal{D})|| \leq ||f|| + \mathcal{O}(\mu) + \mathcal{O}(\mu).$$

For small enough μ, β, this shows that the operator does not expand function values in norm, thus proving non-expansiveness, and i) is demonstrated.

Now let $E_m = f - S_m^{\mu,\beta}(f,\mathcal{D})$, where E_m is the approximation error. The recursion gives

$$||E_{m+1}|| = ||E_m - \beta\langle f_m, \varphi_{m+1}\rangle \ \leq \gamma_m||E_m||.$$

We build the sequence γ_m such as $\gamma_m = u_m(1 + m\beta(2-\beta)\mu^2)^{-\frac{\tau}{2}}$. Summing over iteration, we get:

$$||E_m|| \leq \gamma_m||f||$$

thus, the contraction property holds. □

Lemma 2. *Let the m^{th} minimal worst case error be define as the following form*

$$A_m^{\mu,\beta} = \inf_{f \neq S_m^{\mu,\beta}} \frac{||f - S_m(f,\mathcal{D})||}{||f||^{1-\tau}||f||_{S_1(\mathcal{D})}}, \tag{16}$$

then

1. *The error of identical zero algorithm is given by*
$$A_1^{\mu,\beta} = \inf_{||f||_{H_{n,\gamma} \leq 1}} ||f||_{L_{2,n}} = ||SMAP_n||. \tag{17}$$
2. *In a given dictionary $\mathcal{D}$,*
$$\inf_{\mathcal{D}} A_m^{\mu,\beta} \leq A_1^{\mu,\beta}$$

Proof of Lemma 2. By definition,

$$A_1^{\mu,\beta} = \inf_{f \neq S_1^{\mu,\beta}} \frac{||f - S_1(f,\mathcal{D})||}{||f||^{1-\tau}||f||_{S_1(\mathcal{D})}},$$

and,

$$A_1^{\mu,\beta} = \inf_{S_1(f,\mathcal{D})} \frac{||f - \beta < f_0, \varphi_1 > ||}{||f||^{1-\tau}||f||_{S_1(\mathcal{D})}}.$$

From lemma 1, we have

$$||f - S_1(f,\mathcal{D})|| \leq u_1||f||(1+\beta(2-\beta)\mu^2)^{-\frac{\tau}{2}}, u_1 > 0,$$

Which means that

$$\frac{||f - \beta < f_0, \varphi_1 > ||}{||f||^{1-\tau}} \leq \frac{(||f||u_1(1+\beta(2-\beta)\mu^2)^{\tau}}{\sqrt{u_1(1+\beta(2-\beta)\mu^2)}}$$

And because $||f||_{S_1(\mathcal{D})} := \inf\{M > 0 : f/M \in S_1(\mathcal{D})\}$, we the above expression can be rewritten as follow:

$$\frac{||f - \beta < f_0, \varphi_1 > ||}{||f||^{1-\tau}||f||_{S_1(\mathcal{D})}} \leq \frac{(||f||u_1(1+\beta(2-\beta)\mu^2)^{\tau}}{M\sqrt{u_1(1+\beta(2-\beta)\mu^2)}}.$$

By taking the $\inf_{f \neq S_m^{\mu,\beta}}(||\cdot||)$ in both side, one has

$$A_1^{\mu,\beta} \leq C||f||^{\tau} \leq ||f||_{L_{2,n}}, \text{ with } C = \frac{(u_1(1+\beta(2-\beta)\mu^2)^{\tau}}{M\sqrt{u_1(1+\beta(2-\beta)\mu^2)}}.$$

When we look, for $\inf_{||f||_{H_{n,\gamma}} \leq 1}(||\cdot||)$, we can choose $c = C^{-1} > 0$ big enough such as $cA_1^{\mu,\beta} \geq c||f||^{\tau} \geq ||f||$. Furthermore,

$$C \inf_{f \neq S_1^{\mu,\beta}}(||\cdot||) \leq \inf_{||f||_{H_{n,\gamma}} \leq 1}(||\cdot||) \leq c \inf_{f \neq S_1^{\mu,\beta}}(||\cdot||),$$

which means that both norms are equivalent, et consequently have the same infimum as required for the first part of the lemma.

In any dictionary $\mathcal{D}$, $||f - S_{m+1}(f,\mathcal{D})|| = ||f - S_m(f,\mathcal{D}) - \beta\langle f_m, \varphi_m\rangle||$, therefore $||f - S_{m+1}(f,\mathcal{D})|| \leq ||f - S_m(f,\mathcal{D})|| + \beta\langle f_m, \varphi_{m+1}\rangle$. By dividing both part by $||f||^{1-\tau}||f||_{S_1(\mathcal{D})}$ and taking the infimum we conclude that

$$\inf_{\mathcal{D}} A_m^{\mu,\beta} \leq A_1^{\mu,\beta}.$$

□

4. Proof of the main result

Proof of Theorem 1. Given a Hilbert space $H_{n,\gamma}$, for any $\mathcal{D} \subset H_{n,\gamma}$, we consider $S_m^{\mu,\beta}(f,\mathcal{D})$ as the swarm-based approximation. The goal of the proof is to provide an explicit upper bound for $\phi_m^{\mu,\beta}(\tau, H_{n,\gamma})$. From the non-expansiveness of $S_m^{\mu,\beta}$ from Lemma 1, we recall that $||S_m^{\mu,\beta}(f,\mathcal{D})|| \leq ||f||$ and $||f - S_m^{\mu,\beta}(f,\mathcal{D})|| \leq u_m||f||(1+m\beta(2-\beta)\mu^2)^{-\frac{\tau}{2}}$. When $f \in S_1(\mathcal{D})$, the quasi-norm $||f||_{S_1(\mathcal{D})}$ is such as $||f||^{1-\tau}||f||_{S_1(\mathcal{D})} = ||f||$.

Now let c_0 be a constant threshold such that $0 < c_0 < 1$. When

$$f(X_i) - \varphi(X_i + \alpha_i d_i) = \alpha_i ||g_i||^2 > 0, \tag{18}$$

α_i can be initialize with a large positive value. If $\phi_i \geq c_0$, the calculation stops and α_i is output. Otherwise, we set α_i to $c_1\alpha_i$ with $0 < c_1 < 1$. Note that if $\phi_i \geq c_0$ (see [16, Section 3.1]), then $f(X_i + \alpha_i d_i)$ is very similar to $\varphi(X_i + \alpha_i d_i)$ and α_i can be accepted. In this case, the value of the function $f(X_i)$ will decrease in the direction of $\alpha_i d_i$. Otherwise, the value of α_i will be decremented and the value of ϕ_i will be re-evaluated until (4) is satisfied. Let now $b_m = ||\alpha_m(-\nabla_{X_m})f_m||^2$; $x_m := \alpha_m < f_{m-1}, \phi_m >; m = 1, 2, ...$ and consider the sequence C_m defined as follows:

$$C_0 := ||f||_{S_1(\mathcal{D})}, C_{m+1} := C_m + \beta x_{m+1}.$$

From Lemma 2.,

$$\inf_{f \in S_1(\mathcal{D})} A_m(f,\mathcal{D}) \leq \frac{||SMAP_n||}{||f||_{S_1(\mathcal{D})}} = C_0||SMAP_n||.$$

By taking the infimum on $\mathcal{D}$, we conclude the result. $\square$

5. Numerical analysis

This section presents six common benchmark functions used for evaluating optimization algorithms, particularly swarm-based methods. Each function has unique properties that test the capabilities of optimization algorithms in terms of convergence, exploration, and exploitation.

Rosenbrock Function

$$f(\mathbf{x}) = \sum_{i=1}^{d-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + (1 - x_i)^2\right] \tag{19}$$

Domain: $x_i \in [-5, 10]$
Global Minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (1, \ldots, 1)$
Characteristics: Narrow valley, non-convex, difficult for algorithms to converge.

Rastrigin Function

$$f(\mathbf{x}) = 10d + \sum_{i=1}^{d} \left[x_i^2 - 10\cos(2\pi x_i)\right] \tag{20}$$

Domain: $x_i \in [-5.12, 5.12]$
Global Minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (0, \ldots, 0)$
Characteristics: Highly multimodal, many local minima.

Ackley Function

$$f(\mathbf{x}) = -a \exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(cx_i)\right) + a + \exp(1) \tag{21}$$

Typically, $a = 20$, $b = 0.2$, $c = 2\pi$.

Domain: $x_i \in [-32.768, 32.768]$
Global Minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (0, \ldots, 0)$
Characteristics: Multimodal, large flat region with narrow global minimum.

Griewank Function

$$f(\mathbf{x}) = 1 + \frac{1}{4000}\sum_{i=1}^{d} x_i^2 - \prod_{i=1}^{d}\cos\left(\frac{x_i}{\sqrt{i}}\right) \tag{22}$$

Domain: $x_i \in [-600, 600]$
Global Minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (0, \ldots, 0)$
Characteristics: Many regularly distributed local minima.

Solomon Function

$$r = \sqrt{\sum_{i=1}^{d} x_i^2} \tag{23}$$

$$f(\mathbf{x}) = 1 - \cos(2\pi r) + 0.1r \tag{24}$$

Domain: $x_i \in [-100, 100]$
Global Minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (0, \ldots, 0)$
Characteristics: Radially symmetric, multimodal.

Schwefel Function

$$f(\mathbf{x}) = 418.9829 \times d - \sum_{i=1}^{d} x_i \sin(\sqrt{|x_i|}) \tag{25}$$

Domain: $x_i \in [-500, 500]$
Global Minimum: $f(\mathbf{x}^*) = 0$ at $\mathbf{x}^* = (420.9687, \ldots, 420.9687)$
Characteristics: Many deep local minima, deceptive landscape.

The following table explains the swarm Algorithm Parameters.

- **Swarm size** (m): Larger populations improve the algorithm's ability to explore the search space, but computational cost increases.
- **Acceleration coefficient** (β): Balances exploration and exploitation. High values can lead to rapid convergence but risk premature convergence.
- **Inertia weight** (μ): A dynamic μ often improves performance. Typically, μ decreases over time to shift from exploration to exploitation.

The theoretical error bound associated with the convergence of the swarm algorithm is given by:

$$\text{Bound} = \left(1 + m\beta(2-\beta)\mu^2\right)^{-\frac{\tau}{2}} \tag{26}$$

- τ is a parameter controlling the rate of decay in the error bound.
- Increasing m, β, or μ reduces the error bound (improves theoretical convergence) but may have trade-offs in practice.

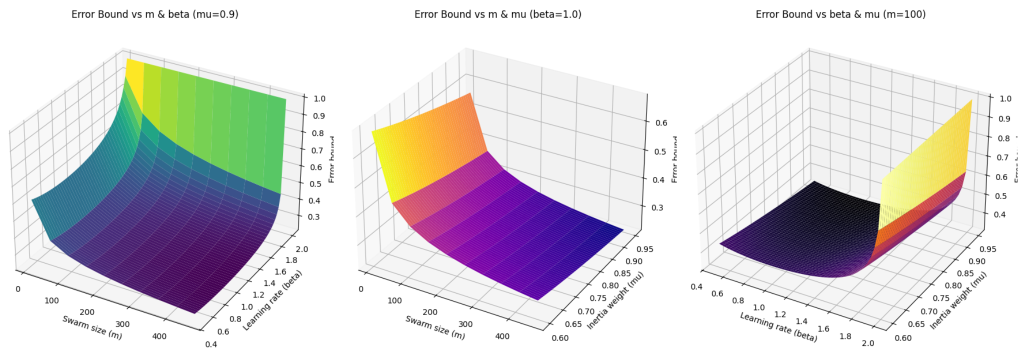

Figure 1. Visualization of the error bound $\left(1 + m\beta(2-\beta)\mu^2\right)^{-\frac{\tau}{2}}$ under different combinations of parameters m, β, and μ. The three plots respectively explore: (1) m and β with $\mu = 0.9$, (2) m and μ with $\beta = 1.0$, and (3) β and μ with $m = 100$.

The three 3D plots visualize the behavior of the error bound:

$$\left(1 + m\beta(2-\beta)\mu^2\right)^{-\frac{\tau}{2}}$$

in relation to the parameters m, β, and μ.

Plot 1: Error Bound vs m and β (fixed $\mu = 0.9$)

- Increasing m leads to a significant reduction in the error bound.
- β should be balanced. Values close to 2 cause $(2-\beta)$ to approach zero, which increases the error bound.

Plot 2: Error Bound vs m and μ (fixed $\beta = 1.0$)

- Larger values of m and μ generally lower the error bound.
- However, very high μ may introduce instability, despite improving convergence rates.

Plot 3: Error Bound vs β and μ (fixed $m = 100$)

- Increasing μ reduces the error bound due to its quadratic effect.
- The choice of β is critical: too low slows convergence, too high increases the error when $(2-\beta)$ becomes too small.

General Insight

- Favor a large m, moderate-to-high μ, and an optimal β typically in the range $[1.2, 1.7]$.
- A careful balance of these parameters ensures fast convergence and algorithm stability.

Step 3: Numerical Examples for $m\beta(2-\beta)\mu^2$

m	β	μ	$2-\beta$	**Expression**	**Value**
20	1.5	0.7	0.5	$20 \times 1.5 \times 0.5 \times 0.7^2$	7.35
30	1.2	0.6	0.8	$30 \times 1.2 \times 0.8 \times 0.6^2$	10.37
40	1.8	0.5	0.2	$40 \times 1.8 \times 0.2 \times 0.5^2$	3.60
25	1.0	0.9	1.0	$25 \times 1.0 \times 1.0 \times 0.9^2$	20.25
50	1.6	0.4	0.4	$50 \times 1.6 \times 0.4 \times 0.4^2$	5.12

Table 1. Computed values for $m\beta(2-\beta)\mu^2$.

Parameter	**Range**	**Impact**	**Tuning Strategy**
m	$[20, 100]$	Linear effect on $m\beta(2-\beta)\mu^2$, decreasing error bound	
β	$(1.0, 2.0)$	Affects $(2-\beta)$: too high reduces exploitation; too low slows convergence	
μ	$(0.4, 0.9)$	Convergence speed via μ^2, higher μ accelerates convergence but may cause instability	
τ	$(0.3, 0.7)$	Controls balance between approximation error and regularity in $\phi_m^{\mu,\beta}$	

Table 2. Recommended tuning strategy for m, β, μ, and τ to minimize the error bound $\phi_m^{\mu,\beta}(\tau, H_{n,\gamma})$.

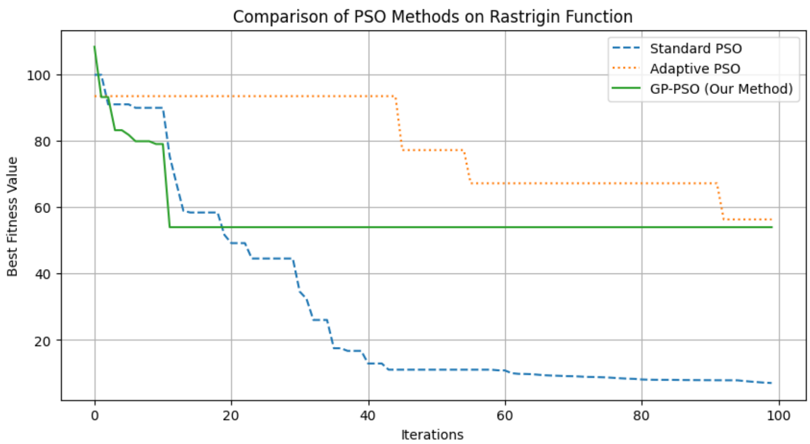

Figure 2. Comparison of PSO Methods on Rastrigin funtion

Step 4: Error Bound Example for $\tau = 0.5$

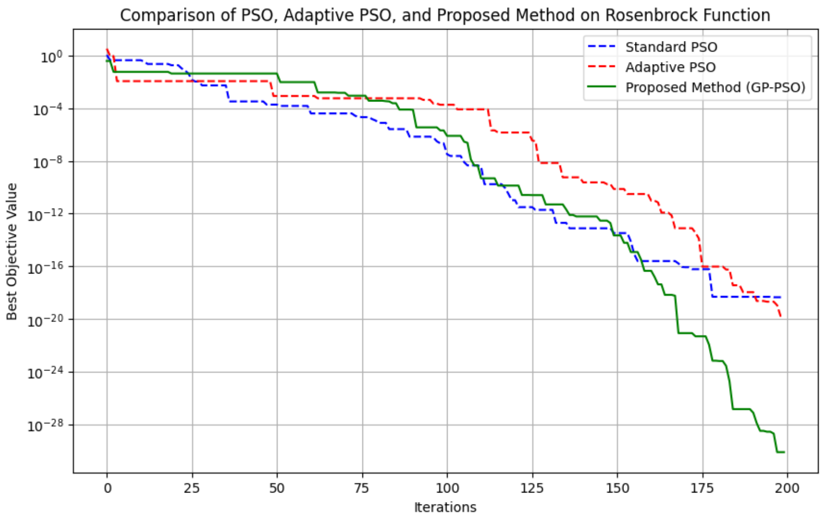

Figure 3. Best Objective value comparison between the Standard PSO, the adaptive PSO and GP-PSO

In Figure 5, the standard PSO (SPSO) show slower convergence before 25 iterations because its get stuck in local optima. The Adaptive PSO show faster convergence and better final objective value more than the SPSO due to dynamic parameters adjustment.

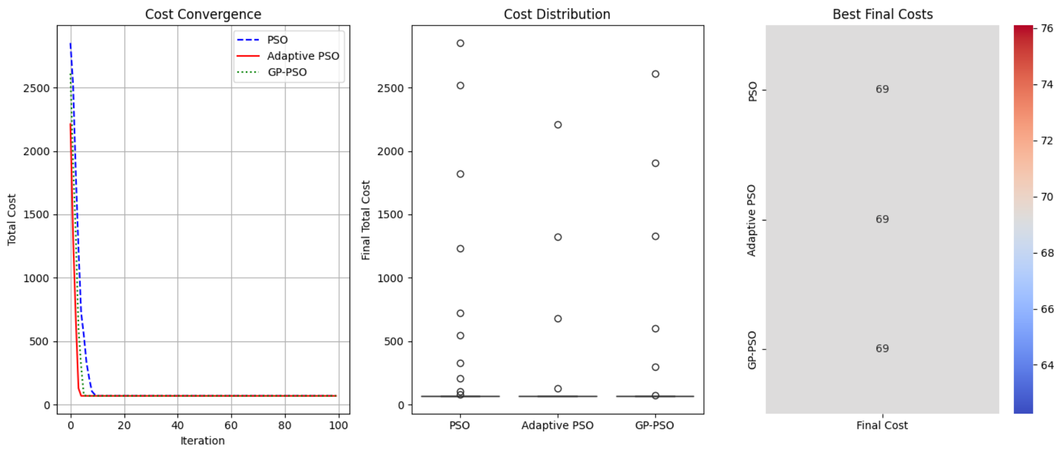

Figure 4. Cost Convergence, Cost Distribution and best final Costs

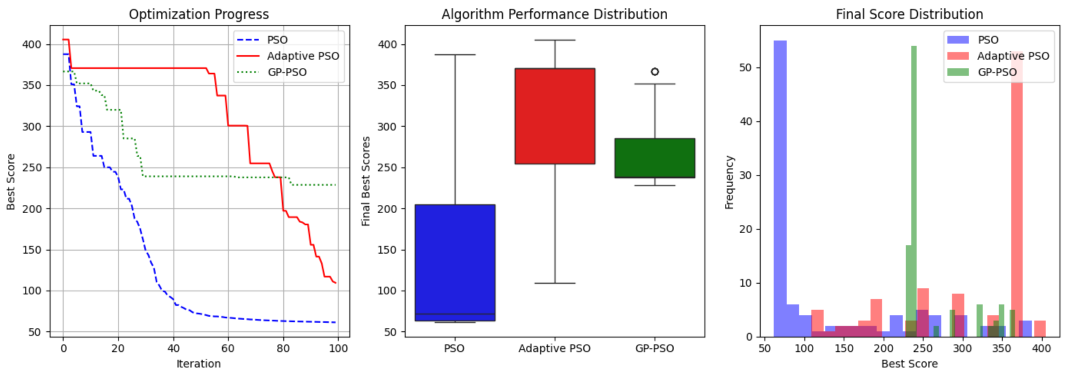

Figure 5. Optimization progress, performance and final score distribution

References

[1] M. Taherkhani and R. Safabakhsh, *A novel stability-based adaptive inertia weight for particle swarm optimization*, Applied Soft Computing, vol. 38, pp. 281–295, 2016.

[2] K. Chen, F. Zhou, and A. Liu, *Chaotic dynamic weight particle swarm optimization for numerical function optimization*, Knowledge-Based Systems, vol. 139, pp. 23–40, 2018

[3] James Stuart Groves, *A study of stochastic processes in Banach spaces* PhD thesis, pp.9-16

[4] Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995. [Google Scholar]

[5] Xionghua Huang, Tiaojun Zeng , and MinSong Li,*A Particle Swarm Optimization Algorithm with Gradient Perturbation and Binary Tree Depth First Search Strategy*, https://doi.org/10.1155/2022/6599899,2022

[6] Temlyakov V. *On the rate of convergence of Greedy Algorithms* Mathematics 2023, 11, 2559. https://doi.org/10.3390/math11112559

[7] J. Lui, Y. Mei, X. Li *An analysis of the Inertia weight parameter for binary Particle Swarm Optimization*, IEEE, 2016

[8] Song, Y.; Lui Y; Chen H?; Dend W. *A Multi-strategy adaptive particle swarm optimization* Electronics 2023.

[9] X. Chen and Y. Li *A modified particle swarm optimization structure resulting in high exploration ability with convergence guaranteed* IEE Trans. Syst. Man Cybern. B Cybern., vol. 37, no. 11, pp. 1591-1604, 2003

[10] Q. Yuan and G. Yin *Analyzing convergence and rates of convergence of particle swarm algortihms using stochastic approximation methods* IEE arXiv:1308.5369v2(2014)

[11] Temlyakov, V.N. Greedy Expansions in Banach spaces. adv compt. Math.2007, 26, 431-449

[12] Chun-Yao Lee, Truong-an Le *An enhanced Binary particle swarm optimizatytion for optimial feature sekection in bearing fault diagnosis of electrical machines* IEE (2021)

[13] M. Berasategui, S. Lassalle, *Weak greedy algorithms and equivalance between semi-greedy and alsmost greedy Markushevich bases* arXiv:2004.06849v4 (2021)

[14] A.A. Khartov, I.A. Limar, *Asymptotic analysis in multivariate worst case approximation with Gaussian kernels,* Journal of Complexity, Volume 82, 2024, 101838, ISSN 0885-064X, https://doi.org/10.1016/j.jco.2024.101838.