

Metodi ibridi di ottimizzazione: algoritmi metaeuristici seguiti da metodi basati sul gradiente

Beatrice Laureti

Università degli studi di Ferrara

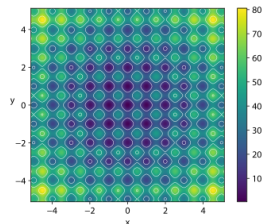
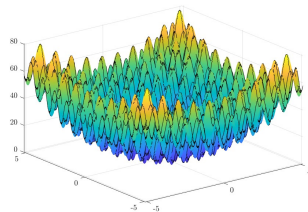
12 febbraio 2026

Obiettivi

- Il progetto si basa sullo sviluppo di algoritmi ibridi che combinano metodi metaeuristici con metodi basati sul gradiente.
- I metodi metaeuristici vengono utilizzati come fase di pre-inizilizzazione, con l'obiettivo di individuare regioni favorevoli nello spazio delle soluzioni da cui avviare l'ottimizzazione locale con i metodi basati sul gradiente.
- Vengono implementate tre strategie ibride:
 - PSO + SGD
 - CBO + SGD
 - PSO + Adam

Funzione di test: Rastrigin

- I tre metodi ibridi vengono applicati a una nota funzione test per valutarne le performance: la funzione Rastrigin.
- Rastrigin è non convessa e viene comunemente utilizzata per testare algoritmi che devono trovare il minimo di una funzione poichè presenta diversi minimi locali, ma un solo minimo globale che si trova in 0 con $f(0) = 0$.
- Rastrigin presenta una superficie altamente ondulata con numerosi minimi locali molto ampi, in cui è facile che l'algoritmo rimanga intrappolato.



Algoritmi metaeuristici

Funzionamento del Particle Swarm Optimization (PSO)

- Inizializzazione di N particelle aventi (X_i^0, V_i^0) posizione e velocità casuali.
- Valutazione della *fitness* $f(X_i^n)$.
- Aggiornamento del best locale (Y_i^n) e del best globale (\bar{Y}^n) .
- Aggiornamento di velocità e posizione:

$$\begin{aligned} X_i^{n+1} &= X_i^n + V_i^{n+1} \\ V_i^{n+1} &= \underbrace{wV_i^n}_{\text{Inerzia}} + \underbrace{c_1 R_{1,i}^n (Y_i^n - X_i^n)}_{\text{Allineamento locale}} + \underbrace{c_2 R_{2,i}^n (\bar{Y}^n - X_i^n)}_{\text{Allineamento globale}} \end{aligned}$$

- w : peso di inerzia.
- c_1, c_2 : coefficienti di accelerazione.
- $R_{1,i}^n, R_{2,i}^n$: matrici diagonali con numeri in $[0, 1]$.

Consensus Based Optimization (CBO)

- Inizializzazione di N particelle con posizioni X_i^0 casuali.
- Stima del minimo globale (punto di consenso) tramite media pesata (distribuzione di Gibbs):

$$\bar{X}_\alpha^n = \frac{\sum_{i=1}^N X_i^n e^{-\alpha f(X_i^n)}}{\sum_{i=1}^N e^{-\alpha f(X_i^n)}}$$

- Aggiornamento delle posizioni:

$$X_i^{n+1} = X_i^n - \underbrace{\Delta t \lambda (X_i^n - \bar{X}_\alpha^n)}_{\text{Allineamento}} + \underbrace{\sqrt{\Delta t} \sigma D(X_i^n - \bar{X}_\alpha^n) \xi_i^n}_{\text{Esplorazione stocastica}}$$

- λ, σ : parametri di consenso e di esplorazione.
- α : parametro di intensità di selezione.
- $D(\cdot)$: matrice di diffusione (esplorazione isotropa o anisotropa).
- ξ_i^n : variabili casuali estratte da una normale standard $N(0, 1)$.

Algoritmi basati sul gradiente

Random Coordinate Descent (RCD)

- **RCD:** Variante dello SGD in cui l'aggiornamento avviene solo su un sottoinsieme casuale D_k delle componenti del punto x .
- A differenza di SGD, la stocasticità non deriva dai campioni scelti ma dalla selezione delle coordinate da aggiornare.
- Il punto viene aggiornato usando il gradiente solo rispetto a un sottoinsieme casuale di componenti, mentre le altre restano invariate. Per ogni coordinata j scelta casualmente in D_k :

$$x_j^{k+1} = x_j^k - \eta_k \partial_j f(x^k)$$

- Passo decrescente $\eta_k = \eta_0/k$.

Adaptive Moment Estimation (Adam)

- Calcolo del momento primo m_k (stima della media dei gradienti) e momento secondo v_k (stima della varianza):

$$m_k = \beta_1 m_{k-1} + (1 - \beta_1) g_k, \quad v_k = \beta_2 v_{k-1} + (1 - \beta_2) g_k^{\odot 2}$$

- Correzione Bias poichè m_0, v_0 partono da zero:

$$\hat{m}_k = \frac{m_k}{1 - \beta_1^k}, \quad \hat{v}_k = \frac{v_k}{1 - \beta_2^k}$$

Per ogni coordinata j selezionata casualmente in D_k :

$$x_j^{k+1} = x_j^k - \eta \frac{\hat{m}_{k,j}}{\sqrt{\hat{v}_{k,j} + \varepsilon}}$$

- β_1, β_2 : coefficienti di decadimento per i momenti.
- η : learning rate (scalato automaticamente per ogni coordinata).
- ε : costante per garantire stabilità numerica.

Implementazione

Implementazione metodi

- Lo studio è stato condotto per $\text{dim} \in \{2, 10, 50\}$, adattando le configurazioni per rispondere alla crescente complessità dello spazio di ricerca.
- Per garantire un confronto equo, i metodi basati su popolazione (PSO, CBO) e quelli basati sul gradiente (SGD, Adam) condividono rispettivamente lo stesso numero di particelle, le iterazioni massime e le medesime condizioni di arresto.
- All'aumentare della dimensionalità, sono stati incrementati il numero di particelle n e delle iterazioni n_{it} .

	PSO e CBO			SGD e Adam		
N	n	n_{it}	$n_{it_{sm}}$	η	n_{it}	$n_{it_{sm}}$
2	50	100	40	$10^{-4} - 10^{-2}$	200	50
10	300	200	40	$10^{-4} - 10^{-2}$	300	50
50	3000	500	40	$10^{-4} - 10^{-2}$	500	50

Parametri utilizzati per ciascun metodo

- Il successo dell'ottimizzazione dipende criticamente dalla scelta degli iperparametri. I valori riportati (fissi e variabili) derivano da test empirici preliminari.
- I parametri espressi come intervalli sono stati scelti come **parametri chiave** per osservare come la loro calibrazione influenzi la convergenza dei metodi.

PSO: $w = 0.4 - 1$, $c_1 = c_2 = 1.5$

CBO: $\alpha = 100$, $\lambda = 0.2 - 0.5$, $\sigma = 1.0 - 2.0$, $\Delta t = 0.1$

SGD: $\eta = 10^{-4} - 10^{-2}$, $|D_k| = \lceil \text{dim}/5 \rceil$

Adam:

$\eta = 10^{-3}$, $|D_k| = \lceil \text{dim}/5 \rceil$, $\beta_1 = 0.5 - 0.9$, $\beta_2 = 0.8 - 0.999$, $\varepsilon = 10^{-8}$.

Criteri di terminazione

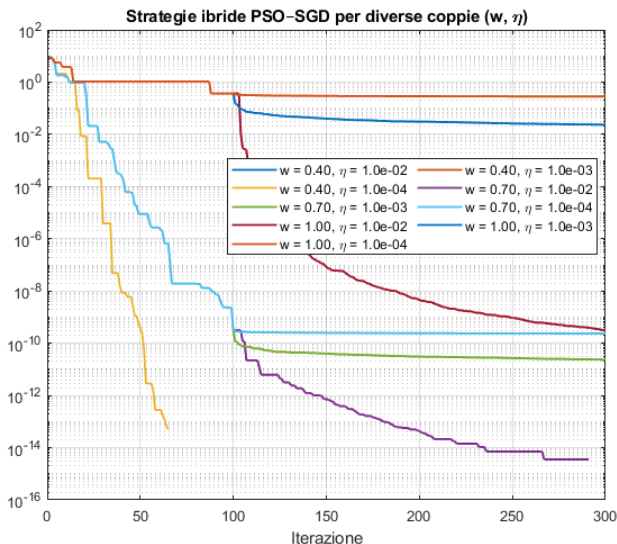
- Tutti i metodi considerati condividono i seguenti criteri di arresto:
 - Raggiungimento del numero massimo di iterazioni n_{it} ;
 - Superamento del numero massimo di iterazioni consecutive senza miglioramento $n_{it_{sm}}$, che indica stagnazione nella ricerca.
- I metodi metaeuristici hanno un ulteriore criterio di terminazione legato alla dinamica collettiva delle particelle, esso prevede il collasso della popolazione:
 - PSO: convergenza delle velocità, quando la norma massima delle velocità delle particelle scende al di sotto di una soglia prefissata.
 - CBO: la dispersione massima delle particelle rispetto al punto di consenso scende sotto una certa soglia. Indica che le particelle hanno raggiunto un consenso.

Risultati ottenuti

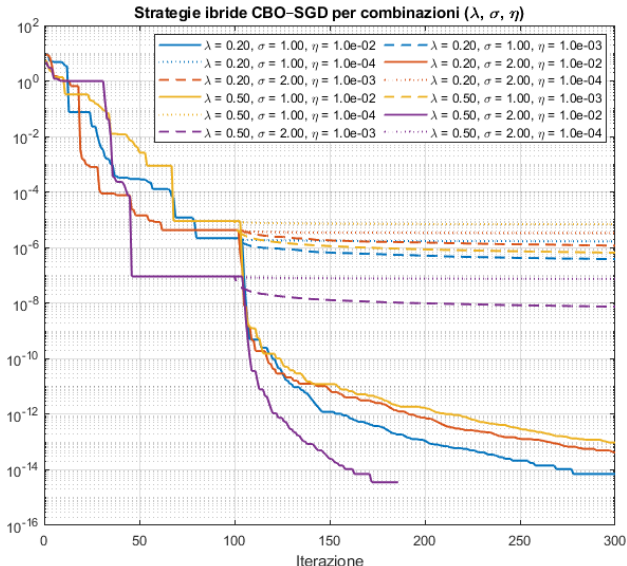
Analisi al variare di alcuni parametri chiave

- **Obiettivo:** Analizzare singolarmente ogni metodo per individuare la configurazione ottimale e procedere successivamente al confronto globale.
- Lo studio si è concentrato sulla variazione dei seguenti parametri chiave:
 - **PSO:** Peso d'inerzia w
 - **CBO:** Consenso λ e Rumore σ
 - **SGD:** Learning rate η
 - **Adam:** Coefficienti di decadimento β_1, β_2
- Per garantire la chiarezza visiva e la confrontabilità dei grafici:
 - In questa fase: Unico criterio di arresto basato sulle iterazioni massime.
 - Fase successiva: Introduzione dei criteri di terminazione dinamici per il confronto congiunto delle performance.

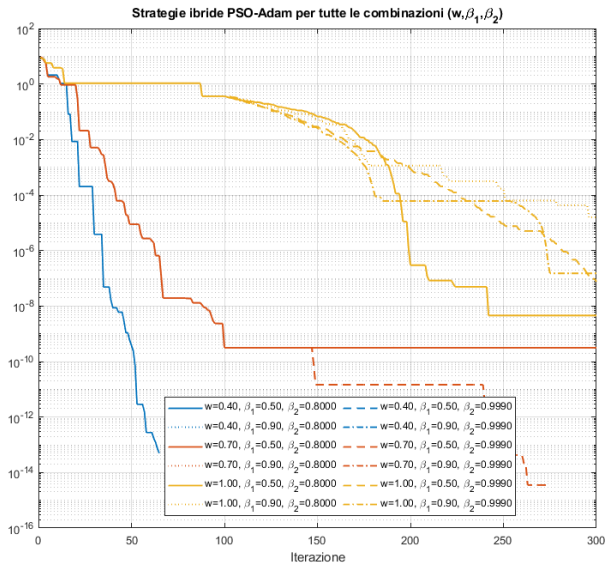
$N=2$



N=2



N=2



Risultati in dimensione $N = 2$

- **PSO:**

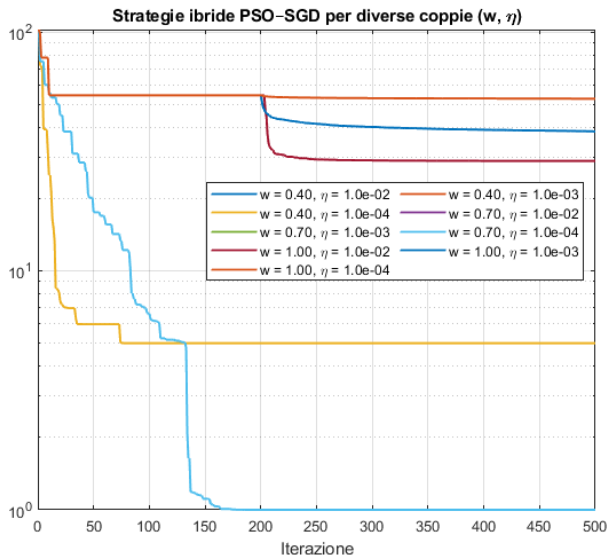
- Individua il minimo globale già nelle prime iterazioni.
- Precisione di macchina raggiunta (nella configurazione migliore), fase di gradiente superflua.

- **CBO:**

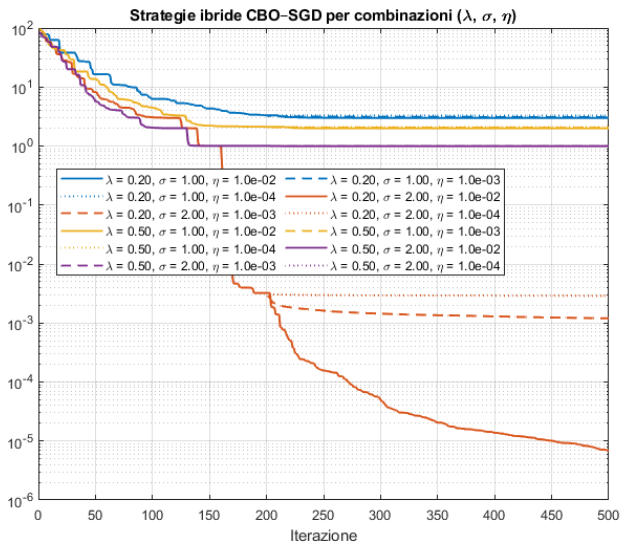
- Convergenza globale garantita in tutte le configurazioni.
- SGD migliora significativamente l'accuratezza finale.

- La soluzione di PSO è più accurata di quella CBO.

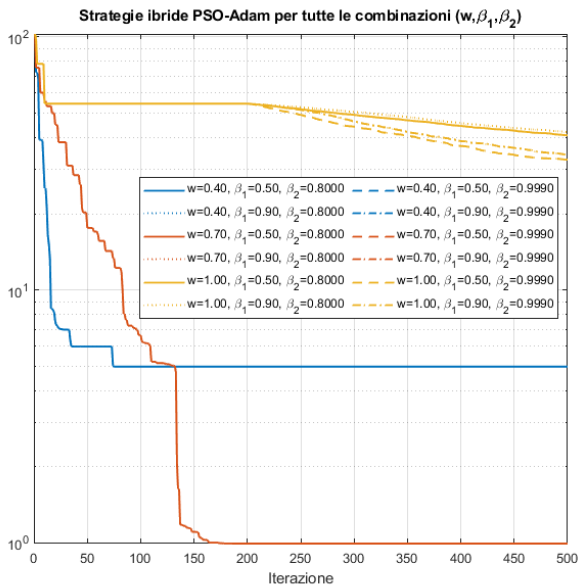
$N=10$



N=10



N=10



Risultati in dimensione $N = 10$

- **PSO:**

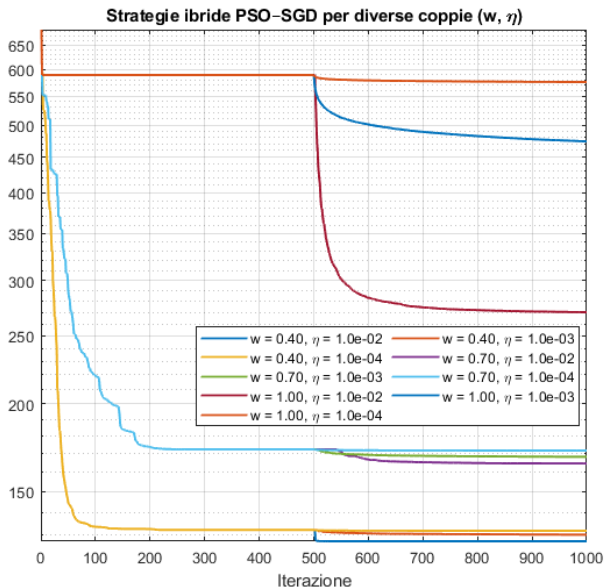
- Tendenza a stabilizzarsi in minimi locali (seppur vicini al globale).
- SGD/Adam non apportano miglioramenti significativi.

- **CBO:**

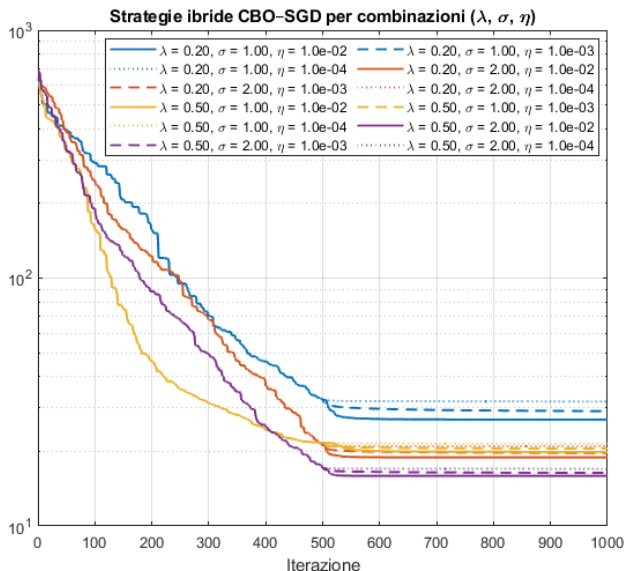
- Maggiore esplorazione dello spazio con raggiungimento del minimo globale.
- SGD apporta un affinamento significativo della soluzione.

- All'aumentare della dimensionalità, la capacità di esplorazione stocastica di CBO diventa determinante rispetto a PSO.

N=50

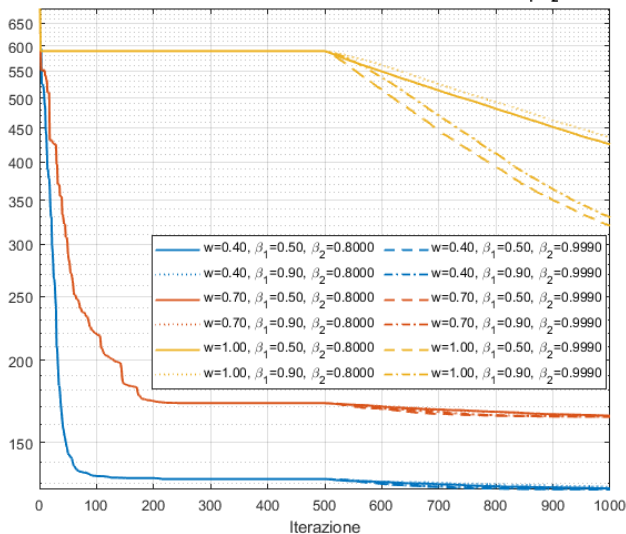


$N=50$



N=50

Strategie ibride PSO-Adam per tutte le combinazioni (w, β_1, β_2)



Risultati in dimensione $N = 50$

- Nonostante l'incremento di particelle e iterazioni, né PSO né CBO raggiungono il minimo globale.

Entrambi i metodi rimangono bloccati in minimi locali con valori di $f(x) \in [10^1, 10^2]$.

- SGD e Adam effettuano un raffinamento locale, ma non possono correggere l'errore di inizializzazione dovuto alla metaeuristica.
- All'aumentare di dim , lo spazio di ricerca cresce esponenzialmente: il panorama della funzione Rastrigin diventa troppo complesso per i parametri impostati, rendendo il minimo globale irraggiungibile per questi algoritmi.

Considerazioni generali sui parametri

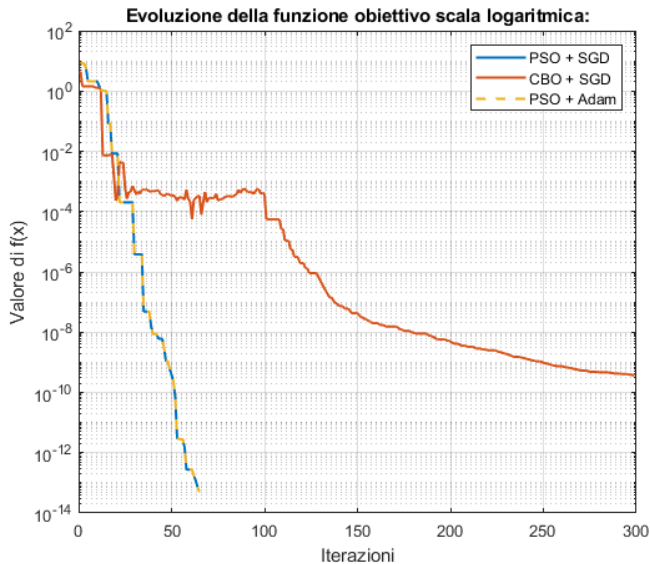
Dall'analisi condotta sulle diverse dimensioni del problema emergono i seguenti set di parametri ottimali:

- **PSO (w):** Valori di inerzia contenuti ($0.4 - 0.7$) risultano preferibili. Inerzie troppo elevate ritardano eccessivamente la convergenza a favore di un'esplorazione spesso inconcludente.
- **CBO (σ, λ):** Sono preferibili σ elevati per esplorare maggiormente lo spazio. $\lambda \in [0.2, 0.5]$ bilancia correttamente la convergenza verso il punto di minimo e la diversità della popolazione.
- **SGD (η):** Il learning rate $\eta = 10^{-2}$ è il più robusto. Valori inferiori ($10^{-3}, 10^{-4}$) tendono a rallentare eccessivamente la dinamica, portando la particella a rimanere bloccata lontano dal minimo.
- **Adam (β_1, β_2):** Performance migliori con valori elevati e la condizione $\beta_1 < \beta_2$.

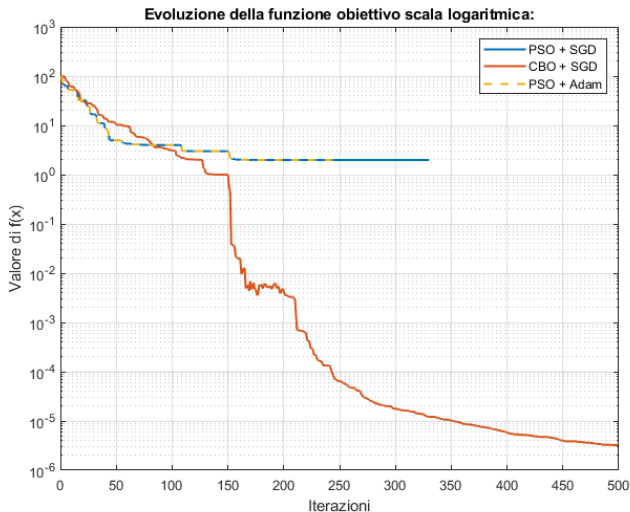
Analisi comparativa degli algoritmi ibridi

- **Obiettivo:** Valutare le performance dei tre algoritmi ibridi proposti, mettendoli a confronto diretto in uno scenario ottimizzato.
- Utilizzo dei parametri migliori individuati per ciascun metodo nell'analisi preliminare e mantenimento dei parametri statici definiti nel setup iniziale:
 - N=2: $w = 0.4, \lambda = 0.5, \sigma = 2, \eta = 10^{-2}, \beta_1 = 0.9, \beta_2 = 0.999$.
 - N=10: $w = 0.7, \lambda = 0.2, \sigma = 2, \eta = 10^{-2}, \beta_1 = 0.9, \beta_2 = 0.999$.
 - N=50: $w = 0.4, \lambda = 0.5, \sigma = 2, \eta = 10^{-2}, \beta_1 = 0.9, \beta_2 = 0.999$.
- Reintroduzione di tutti i criteri di arresto.

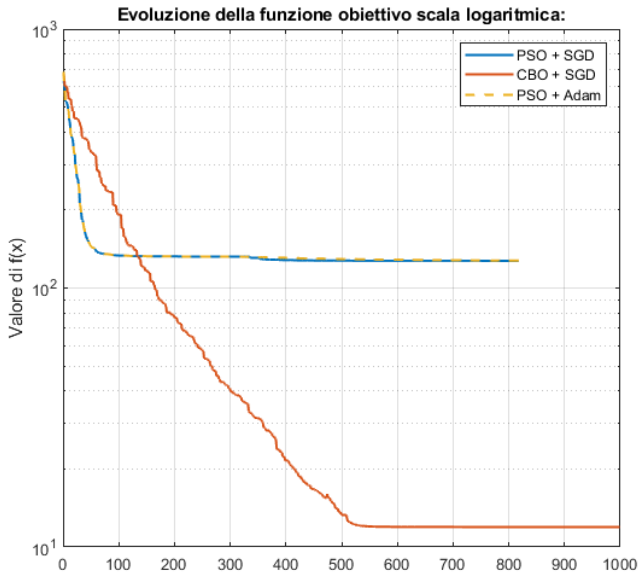
$N=2$



$N=10$



$N=50$



Analisi della fase metaeuristica (esplorazione)

- Nonostante l'elevato costo computazionale, la fase metaeuristica è essenziale per l'esplorazione globale dello spazio delle soluzioni.
- **PSO vs CBO:**
 - $N = 2$: **PSO** è più rapido di CBO nell'individuare il minimo globale, con convergenza veloce già nelle prime iterazioni.
 - $N = 10$: **CBO** mostra capacità esplorative superiori. Mentre PSO rimane intrappolato in minimi locali, CBO riesce a entrare nel bacino del minimo globale.
 - $N = 50$: Entrambi mostrano limiti sulla funzione di Rastrigin. Essi rimangono intrappolati in minimi locali (ordine 10^1 – 10^2), sebbene CBO fornisca soluzioni qualitativamente migliori.

Analisi dei metodi ibridi e risultati globali

- Il gradiente agisce come raffinamento di precisione, ma non può compensare una fase esplorativa fallimentare, dipendendo fortemente dalla soluzione trovata nella fase metaeuristica.
- **Confronto tra algoritmi ibridi:**
 - $N = 2$: **PSO-SGD/Adam** risultano preferibili, poiché CBO mantiene una dinamica esplorativa eccessiva anche dopo aver trovato il minimo.
 - $N \geq 10$: **CBO-SGD** è il più efficace grazie alla maggiore capacità esplorativa di CBO.
- **SGD vs Adam:**
 - **SGD** risulta preferibile su Rastrigin: maggiore capacità esplorativa e minor costo computazionale.
 - **Adam** tende a stabilizzare troppo gli aggiornamenti (stima dei momenti), portando a una convergenza prematura nei minimi locali della funzione.