

Método Dialético de Otimização usando o Princípio da Máxima Entropia

Wellington Pinheiro dos Santos¹ e Francisco Marcos de Assis²

¹Escola Politécnica de Pernambuco
Universidade de Pernambuco
50.720-001, Recife, Pernambuco
e-mail: wellington.santos@ieee.org

²Departamento de Engenharia Elétrica
Centro de Engenharia Elétrica e Informática
Universidade Federal de Campina Grande
58.109-970, Campina Grande, Paraíba
e-mail: fmarcos@dee.ufcg.edu.br

Resumo - A Biologia, a Psicologia e as Ciências Sociais estão intrinsecamente ligadas aos fundamentos do desenvolvimento de métodos e algoritmos em computação evolucionária, como fica evidente em abordagens como algoritmos genéticos, programação evolucionária e otimização por enxame de partículas. Contudo, a Filosofia ainda é vista como algo enigmático, apesar da natureza sistemática de métodos investigativos como a dialética. Neste trabalho é proposto um novo método de otimização baseado na dialética como definida pelos trabalhos de Hegel e pela Filosofia da Práxis, utilizando funções de pertinência obtidas da aplicação do princípio da máxima entropia adaptado a entropia *fuzzy*, para modelar as interações entre os pólos, as unidades básicas que compõem sistemas dialéticos. Para validar a proposta foram feitas diversas simulações usando funções de teste conhecidas. Este trabalho provou que o algoritmo dialético proposto pode atingir bons resultados em problemas de otimização local e global.

Palavras-chave - Métodos de otimização, Dialética, Princípio da máxima entropia, Entropia *fuzzy*

1. Introdução

A concepção dialética da realidade é um tipo de método filosófico investigativo para analisar processos presentes na natureza e em sociedades humanas. Suas origens estão conectadas às filosofias das antigas civilizações da Grécia, da China e da Índia e estão fortemente relacionadas ao pensamento de Heráclito, Platão, e às filosofias de religiões como o Confucionismo, o Budismo e a Escola Zen [1]. Como um método geral de análise, a dialética experimentou um progresso considerável devido ao desenvolvimento da filosofia alemã do século XIX, com a dialética de Hegel, e do começo do século XX, com os trabalhos de Marx, Engels, Gramsci e Lukács [5].

A maioria dos paradigmas de Inteligência Computacional, particularmente redes neurais, computação evolucionária e algoritmos de inspiração sócio-cultural (comportamental), são baseados em algum tipo de teoria voltada a aplicações gerais, embora na prática sejam teorias bastante incompletas. As abordagens baseadas em redes neurais se apóiam sobre determinados modelos do cérebro. A computação evolucionária é baseada na Teoria da Evolução [3]; por sua vez, algoritmos de inspiração sócio-cultural são baseados no estudo de populações, tais como os algoritmos de formigas e em otimização por enxame de partículas (*Particle Swarm Optimization*, PSO) [3]. O Método Dialético Objetivo (*Objective Dialectical Method*, ODM) é um método baseado na dialética para as tarefas de busca e otimização. Neste trabalho se busca apresentar o ODM como uma proposta de algoritmo dialético para busca e otimização, aproveitando as formalizações da Computação Evolucionária e da Lógica Fuzzy.

Também se busca mostrar que é possível usar o Princípio da Máxima Entropia para se obter as expressões das funções de pertinência empregadas no algoritmo proposto, uma vez que a maximização da entropia assim definida tanto garante a convergência do algoritmo quanto garante a máxima exploração da vizinhança do ponto ótimo.

2. Entropia Fuzzy

A necessidade de representar conhecimento disponível de variáveis cujos valores são desconhecidos tem se tornado uma necessidade crescente em sistemas de informação [9]. Existem diversos tipos de medida de incerteza. Pode-se afirmar que medidas *fuzzy* podem representar incertezas probabilísticas (probabilidades) e possibilísticas, sendo possível construir diversas formas de medida de entropia [9][4]. Seja V uma variável que pode assumir os n estados possíveis:

$$X = \{x_1, x_2, \dots, x_n\},$$

onde X é o conjunto de estados, uma medida *fuzzy* μ é um mapeamento de um subconjunto de X no intervalo $[0,1]$, ou seja, $\mu : 2^X \rightarrow [0, 1]$, obedecendo às seguintes propriedades [9]:

1. $\mu(X) = 1$;
2. $\mu(\emptyset) = 0$;
3. $A \subseteq B \Rightarrow \mu(A) \leq \mu(B), \forall A, B \subseteq 2^X$.

No caso de se ter a incerteza expressa na forma probabilística, o conhecimento a respeito de uma variável desconhecida pode ser representado por uma medida *fuzzy* μ onde $\mu(\{x_i\}) = p_i$ e conseqüentemente [9]:

$$\mu(E) = \sum_{x_i \in E} \mu(\{x_i\}),$$

onde $\mu(E)$ nesta situação acaba sendo uma medida de probabilidade de E . Uma vez que $\mu(X) = 1$, tem-se:

$$\mu(X) = \sum_{x_i \in X} \mu(\{x_i\}) = \sum_{i=1}^n p_i = 1.$$

Uma definição possibilística da incerteza do evento E é a que segue:

$$\mu(E) = \max_{x_i \in E} \mu(\{x_i\}).$$

Em diversas aplicações medidas não aditivas são necessárias para expressar incerteza, tais como as medidas *fuzzy* são medidas não aditivas por excelência, dado que obedecem às três condições citadas nos parágrafos anteriores, ou seja, $\mu(X) = 1$, $\mu(\emptyset) = 0$ e $A \subseteq B \Rightarrow \mu(A) \leq \mu(B), \forall A, B \subseteq 2^X$, e não respeitam a aditividade, uma vez que, caso $\mu(A \cup B) \neq \mu(A) + \mu(B)$, tem-se $A \cap B = \emptyset$ [9]. Na verdade, é possível também ter-se a seguinte definição: $\mu(A \cup B) = \max\{\mu(A), \mu(B)\}$.

Assim, para medir a quantidade de informação média *fuzzy*, é preciso incluir também a informação complementar, como é o caso da seguinte entropia obtida a partir da generalização de um sistema com dois estados [4]:

$$H(A) = - \sum_{i=1}^n [h_i(A) + \bar{h}_i(A)],$$

para

$$h_i(A) = \mu_A(x_i) \ln \mu_A(x_i),$$

e

$$\bar{h}_i(A) = (1 - \mu_A(x_i)) \ln(1 - \mu_A(x_i)),$$

onde $\mu_A(x_i)$ é o grau de pertinência de x_i ao conjunto A . Contudo, a tentativa de se aplicar o Princípio da Máxima Entropia usando as entropias apresentadas por Fan e Ma (2002) resulta em equações do segundo grau que podem não ter solução real. Assim, Yager (2000) propõe uma fórmula de entropia *fuzzy* similar à Entropia de Shannon, definida para a medida *fuzzy* μ da forma que segue [9]:

$$H(\mu) = - \sum_{i=1}^n V_i \ln V_i, \tag{Eq. 1}$$

onde V_i é o índice de Shapley, uma medida de incerteza que inclui os valores de pertinência μ e seus complementos, definido por [9]:

$$V_i = \sum_{k=0}^{n-1} \gamma_k \sum_{E \subseteq F_i, |E|=k} [\mu(E \cup \{x_i\}) - \mu(E)], \tag{Eq. 2}$$

onde $F_i = X - \{x_i\}$, $|E|$ é a cardinalidade do conjunto E , e

$$\gamma_k = \frac{k!(n-k-1)!}{n!}, \tag{Eq. 3}$$

onde $V_i \in [0, 1]$ e $\sum_{i=1}^n V_i = 1$.

Em problemas de agrupamento (*clustering*), as variáveis envolvidas assumem apenas um único estado, dado que agora não faz sentido afirmar que um objeto pertence igualmente a dois grupos (*clusters*) ao mesmo tempo, por exemplo; assim, as medidas de incerteza *fuzzy* passam a ser medidas probabilísticas. Logo [9]:

$$\mu(E \cup \{x_i\}) - \mu(E) = \mu(\{x_i\}) = p_i,$$

o que resulta

$$V_i = \sum_{k=0}^{n-1} \gamma_k \sum_{E \subseteq F_i, |E|=k} p_i = p_i \sum_{k=0}^{n-1} \gamma_k \sum_{E \subseteq F_i, |E|=k} 1.$$

Contudo, para qualquer valor de k , existem exatos $\frac{(n-1)!}{k!(n-k-1)!}$ subconjuntos E tais que $E \subseteq X - \{x_i\}$ e $|E| = k$, o que resulta [9]:

$$\sum_{E \subseteq F_i, |E|=k} 1 = \frac{(n-1)!}{k!(n-k-1)!},$$

de onde vem que

$$V_i = p_i \sum_{k=0}^{n-1} \gamma_k \frac{(n-1)!}{k!(n-k-1)!}.$$

Assim,

$$V_i = p_i \sum_{k=0}^{n-1} \frac{(n-1)!}{n!} = p_i \sum_{k=0}^{n-1} \frac{1}{n} = p_i,$$

de onde se conclui que, para casos onde a medida de incerteza *fuzzy* coincide com uma medida probabilística, ou seja, $\mu_i = \mu(\{x_i\}) = p_i$, como no caso de problemas de agrupamento, a Entropia de Shapley coincide com a Entropia de Shannon [9]:

$$H(\mu) = - \sum_{i=1}^n V_i \ln V_i = - \sum_{i=1}^n p_i \ln p_i. \quad \text{Eq. 4}$$

3. Otimização por Enxame de Partículas

Os algoritmos de Otimização por Enxame de Partículas (*Particle Swarm Optimization*, PSO) foram criados por James Kennedy e Russel Eberhart, em 1995, respectivamente um psicólogo social e um engenheiro eletricitista [6]. Os PSOs são baseados no comportamento e movimento de bandos de animais, tais como peixes e pássaros, sendo portanto algoritmos baseados em teorias que descrevem comportamento social animal, possuindo elementos em comum com os algoritmos genéticos e com programação evolucionária [2][6].

De forma semelhante aos algoritmos genéticos, o PSO é inicializado com uma população inicial aleatória. Entretanto, enquanto nos algoritmos genéticos os indivíduos dessa população inicial são representados por cromossomos, no PSO estão associados a cada indivíduo um vetor posição e um vetor de velocidades. Além disso, no PSO não ocorrem mutações nem seleção de indivíduos. Assim, a cada iteração, apenas são ajustadas posições e velocidades dos diferentes indivíduos na direção da *melhor posição global* e da *melhor posição individual*, de acordo com uma determinada *função objetivo*, segundo a seguinte expressão canônica [3]:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1), \quad \text{Eq. 5}$$

desde que

$$\mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + \Delta\mathbf{v}_1(t) + \Delta\mathbf{v}_2(t), \quad \text{Eq. 6}$$

onde

$$\Delta\mathbf{v}_1(t) = c_1 r_1(t)(\mathbf{p}_i(t) - \mathbf{x}_i(t)),$$

$$\Delta\mathbf{v}_2(t) = c_2 r_2(t)(\mathbf{p}_g(t) - \mathbf{x}_i(t)),$$

para $1 \leq i \leq m$, onde m é o número de partículas do enxame; w é o fator de inércia, onde $0 < w < 1$; $r_1(t)$ e $r_2(t)$ são números aleatórios uniformemente distribuídos no intervalo $[0,1]$; c_1 e c_2 são constantes de constrição, também chamados de

coeficientes de aceleração, de forma que $c_1 + c_2 = 4$ (tipicamente, $c_1 = 2 + \Delta$ e $c_2 = 2 - \Delta$, onde $\Delta \approx 0$), sendo que c_1 é o peso devido à consciência própria da partícula, consciência individual ou consciência local, dependendo da implementação, enquanto c_2 é o peso devido à consciência global; \mathbf{x}_i é a posição, enquanto \mathbf{v}_i é a velocidade da i -ésima partícula; \mathbf{p}_g é a melhor posição global, ou seja

$$f(\mathbf{p}_g) = \max_{1 \leq j \leq m} f(\mathbf{x}_j), \quad \text{Eq. 7}$$

enquanto \mathbf{p}_i é a melhor posição individual ou local em relação à i -ésima partícula. Caso a implementação seja baseada na melhor posição individual, a definição utilizada será a que segue:

$$f(\mathbf{p}_i(t)) = \max_{0 \leq t' \leq t} f(\mathbf{x}_i(t')). \quad \text{Eq. 8}$$

Como não possui mecanismos de mutação e seleção de indivíduos, o PSO é uma classe de algoritmos relativamente simples de serem implementados, já que se reduzem a um processo iterativo aliado à busca dos melhores indivíduos por meio da avaliação da função objetivo em todos os indivíduos (partículas). Diversos resultados mostram que, para uma classe relativamente grande de problemas, o PSO atinge bons resultados [3].

4. Método Dialético de Otimização

Os conceitos básicos da dialética foram definidos matematicamente da forma que segue:

Pólo: É a menor unidade integrante de um sistema dialético. Dado um conjunto de pólos $\Omega = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$, um pólo i está associado a um vetor de pesos $\mathbf{w}_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n})^T$, onde $\mathbf{w}_i \in S$, m é o número de pólos e n é a dimensionalidade do sistema. Um pólo também executa o papel de um candidato à solução. A dimensionalidade do problema coincide com a dimensionalidade da função objetivo $f : S \rightarrow \mathbb{R}$ que se deseja maximizar, onde $S \subseteq \mathbb{R}^n$ e $\Omega \subseteq S$.

Força social: A cada pólo i está associada uma *força social* cujo cálculo coincide com a avaliação da função objetivo f no pólo de ordem i , ou seja, a força social no i -ésimo pólo é dada pelo termo $f(\mathbf{w}_i)$.

Hegemonia: No processo de luta de pólos, diz-se que um pólo k exerce a *hegemonia* no momento t quando:

$$f(\mathbf{w}_k(t)) = f_C(t) = \max_{1 \leq j \leq m(t)} f(\mathbf{w}_j(t)),$$

onde $1 \leq k \leq m(t)$. O vetor $\mathbf{w}_C(t) = \mathbf{w}_k(t)$ é chamado de *pólo hegemônico presente*, ou *pólo hegemônico contemporâneo*, enquanto $f_C(t)$ é a *força hegemônica presente*, ou *força hegemônica contemporânea*. A *força hegemônica histórica* no instante t , $f_H(t)$, é dada por:

$$f_H(t) = \max_{0 \leq t' \leq t} f_C(t'),$$

onde $\mathbf{w}_H(t) = \mathbf{w}_C(t')$, para $f(\mathbf{w}_C(t')) = f_H(t)$ e $0 \leq t' \leq t$.

Antítese absoluta: Dado x tal que $a \leq x \leq b$, onde $a, b \in \mathbb{R}$, o *oposto* de x é dado por [8]:

$$\check{x} = b - x + a.$$

Assumindo $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ e que $\mathbf{x} \in S \Rightarrow r_i \leq x_i \leq s_i, \forall i = 1, 2, \dots, n$, onde r_i e s_i são os limites inferior e superior da i -ésima dimensão de S , o *vetor oposto* associado, $\check{\mathbf{x}} = (\check{x}_1, \check{x}_2, \dots, \check{x}_n)^T$ tem suas coordenadas calculadas da forma que segue [8]:

$$\check{x}_i = s_i - x_i + r_i,$$

onde $i = 1, 2, \dots, n$. O *vetor antítese absoluta* do pólo \mathbf{w} é definido pelo seu vetor oposto $\check{\mathbf{w}}$. Rahnamayan *et al* afirmam que, em programação evolucionária e em otimização por enxame de partículas, a presença de pares opostos na população inicial tipicamente acelera em 10% a convergência dos algoritmos [8].

Contradição: A contradição entre os pólos \mathbf{w}_p e \mathbf{w}_q é definida por:

$$\delta_{p,q} = d(\mathbf{w}_p, \mathbf{w}_q),$$

onde $d : S^2 \rightarrow \mathbb{R}_+$ é uma *função de distância*, de forma que $\delta_{p,q} = \delta_{q,p}, \forall p, q$. Uma função de distância típica é a função de distância euclidiana.

Síntese: De acordo com a concepção dialética, a síntese é a resolução da contradição entre dois pólos, onde um exerce o papel de tese e o outro, de antítese [7]. O pólo $\mathbf{w}_r \in S$ é a síntese entre os pólos \mathbf{w}_p e \mathbf{w}_q se for obtido da forma

$$\mathbf{w}_r = g(\mathbf{w}_p, \mathbf{w}_q),$$

onde $g : S^2 \rightarrow S$. Há duas sínteses possíveis:

$$w_{u,i} = \begin{cases} w_{p,i}, & i \bmod 2 = 0, \\ w_{q,i}, & i \bmod 2 = 1, \end{cases},$$

$$w_{v,i} = \begin{cases} w_{p,i}, & i \bmod 2 = 1, \\ w_{q,i}, & i \bmod 2 = 0, \end{cases},$$

para $i = 1, 2, \dots, n$.

O método dialético objetivo pode ser adaptado para as tarefas de busca e otimização da seguinte maneira:

Definem-se a quantidade inicial $m(0)$ de pólos integrantes do sistema dialético $\Omega(0)$, juntamente com o *número de fases históricas*, n_P , e a *duração de cada fase histórica*, n_H . O número inicial de pólos deve ser par, de forma que metade do número de pólos seja gerada de forma aleatória, dentro do domínio da função objetivo, enquanto a outra metade é obtida pela geração de pólos antítese absoluta. Assim:

$$w_{i,j}(0) = \begin{cases} U(r_j, s_j), & 1 \leq j \leq \frac{1}{2}m(0), \\ \check{w}_{i',j}(0), & 1 + \frac{1}{2}m(0) \leq j \leq m(0), \end{cases} \quad \text{Eq. 9}$$

para $i' = i - \frac{1}{2}m(0)$, $1 \leq i \leq m(0)$ e $1 \leq j \leq n$, onde n é a dimensionalidade do problema de otimização, $U(r_j, s_j)$ é um número aleatório uniformemente distribuído no intervalo $[r_j, s_j]$ e $S = \bigcap_{j=1}^n [r_j, s_j]$, desde que $s_j > r_j$ e $s_j, r_j \in \mathbb{R}$.

Enquanto não se atinge um máximo de n_P fases históricas e a força hegemônica histórica não é maior do que um dado limiar superior de força (estimativa inicial do valor máximo da função objetivo), $f_H(t) < f_{\text{sup}}$ (critério para se considerar o máximo da função objetivo atingido), repetem-se a etapa de *evolução* e, em seguida, a de *crise revolucionária*:

Evolução: Enquanto não se atinge um máximo de n_H iterações e $f_H(t) < f_{\text{sup}}$, os pólos são ajustados segundo a seguinte expressão:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \Delta \mathbf{w}_{C,i}(t) + \Delta \mathbf{w}_{H,i}(t), \quad \text{Eq. 10}$$

para

$$\Delta \mathbf{w}_{C,i}(t) = \eta_C(t)(1 - \mu_{C,i}(t))^2(\mathbf{w}_C(t) - \mathbf{w}_i(t)), \quad \text{Eq. 11}$$

$$\Delta \mathbf{w}_{H,i}(t) = \eta_H(t)(1 - \mu_{H,i}(t))^2(\mathbf{w}_H(t) - \mathbf{w}_i(t)), \quad \text{Eq. 12}$$

$$0 < \eta_C(t) < 1, \quad \text{Eq. 13}$$

$$0 < \eta_H(t) < 1, \quad \text{Eq. 14}$$

onde $\eta_C(0) = \eta_H(0) = \eta_0$ e $0 < \eta_0 < 1$. Os termos $\Delta \mathbf{w}_{C,i}(t)$ e $\Delta \mathbf{w}_{H,i}(t)$ modelam as influências das hegemonias presente e histórica, nesta ordem, sobre o i -ésimo pólo, enquanto $\eta_C(t)$ e $\eta_H(t)$ são os respectivos passos de atualização dos pólos, atualizados a cada iteração e a cada fase histórica, respectivamente, de forma que

$$\eta_C(t+1) = \alpha \eta_C(t), \quad \text{Eq. 15}$$

ao final de cada iteração e

$$\eta_H(t+1) = \alpha\eta_H(t), \quad \text{Eq. 16}$$

ao final de cada fase histórica, para $\alpha < 1$ (tipicamente, $\alpha = 0,9999$).

Quanto à convergência do algoritmo dialético de busca, é necessário que o processo minimize as seguintes quantidades:

$$E_H = \sum_{i=1}^m \mu_{H,i} |f(\mathbf{w}_i) - f_H|, \quad \text{Eq. 17}$$

$$E_C = \sum_{i=1}^m \mu_{C,i} |f(\mathbf{w}_i) - f_C|. \quad \text{Eq. 18}$$

Considerando que na verdade $\mu_{H,i}$ e $\mu_{C,i}$ expressam as probabilidades $p_{H,i}$ e $p_{C,i}$ de o i -ésimo pólo coincidir com o pólo hegemônico histórico e com o pólo hegemônico contemporâneo, respectivamente, pode-se definir uma *entropia histórica* $H(\mu_H)$ e uma *entropia contemporânea* $H(\mu_C)$, que podem ser maximizadas em paralelo, para maximizar a capacidade exploratória do algoritmo:

$$H(\mu_H) = -\sum_{i=1}^m \mu_{H,i} \ln \mu_{H,i}, \quad \text{Eq. 19}$$

$$H(\mu_C) = -\sum_{i=1}^m \mu_{C,i} \ln \mu_{C,i}, \quad \text{Eq. 20}$$

obtendo-se as seguintes expressões de pertinência:

$$\mu_{H,i} = \frac{\exp(-\lambda_H |f(\mathbf{w}_i) - f_H|)}{\sum_{j=1}^m \exp(-\lambda_H |f(\mathbf{w}_j) - f_H|)}, \quad \text{Eq. 21}$$

$$\mu_{C,i} = \frac{\exp(-\lambda_C |f(\mathbf{w}_i) - f_C|)}{\sum_{j=1}^m \exp(-\lambda_C |f(\mathbf{w}_j) - f_C|)}, \quad \text{Eq. 22}$$

onde $\lambda_H > 0$ e $\lambda_C > 0$ são os multiplicadores de Lagrange para as funções de pertinência históricas e contemporâneas, respectivamente. Uma vez que $f_H(t)$ e $f_C(t)$ não são fixos, não é possível determinar algebricamente os multiplicadores de Lagrange $\lambda_H(t)$ e $\lambda_C(t)$. Uma boa escolha poderia ser:

$$\lambda_H(t) = \lambda_C(t) = \frac{1}{m(t)}. \quad \text{Eq. 23}$$

Crise Revolucionária: Na etapa de crise revolucionária são executados os seguintes passos:

Todas as contradições $\delta_{i,j}$ são avaliadas; as contradições menores do que uma *contradição mínima* δ_{\min} implicam na fusão entre os pólos.

A partir das contradições avaliadas na etapa anterior, encontram-se aquelas maiores do que uma *contradição máxima* δ_{\max} ; essas contradições serão consideradas as *contradições principais* do sistema dialético, sendo os pares de pólos envolvidos os *pares tese-antítese*, cujos *pólos síntese* passam também a pertencer ao novo conjunto de pólos.

Adiciona-se o *efeito de crise*, dada a *máxima crise*, χ_{\max} , a todos os pólos do sistema dialético $\Omega(t+1)$, gerando o novo conjunto de pólos, $\Omega(t+2)$, de forma que $\mathbf{w}_k(t+2) \in \Omega(t+2)$, desde que

$$w_{k,i}(t+2) = w_{k,i}(t+1) + \chi_{\max} G(0,1), \quad \text{Eq. 24}$$

para $1 \leq k \leq m(t+1)$ e $1 \leq i \leq n$, onde $G(0,1)$ é um número aleatório de distribuição gaussiana com esperança 0 e variância 1.

Caso o critério de parada ainda não tenha sido atingido, é gerado um novo conjunto de pólos, ampliado por meio da adição dos pólos em antítese antagônica aos pólos já existentes.

5. Resultados Experimentais

O método dialético proposto foi validado pela minimização das seguintes funções de teste [8]:

Modelo esférico:

$$f_1(\mathbf{x}) = \sum_{i=1}^{30} x_i^2,$$

onde $-100 \leq x_i \leq 100$, para $1 \leq i \leq 30$, e $\min(f_1) = f_1(0, \dots, 0) = 0$.

Função de Rosenbrock generalizada:

$$f_5(\mathbf{x}) = \sum_{i=1}^{29} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2],$$

onde $-30 \leq x_i \leq 30$, para $1 \leq i \leq 30$, e $\min(f_5) = f_5(1, \dots, 1) = 0$.

Função de Rastrigin generalizada:

$$f_9(\mathbf{x}) = \sum_{i=1}^{30} [x_i^2 - 10 \cos 2\pi x_i + 10],$$

onde $-5,12 \leq x_i \leq 5,12$, para $1 \leq i \leq 30$, e $\min(f_9) = f_9(0, \dots, 0) = 0$.

Função de Griewank generalizada:

$$f_{11}(\mathbf{x}) = \frac{1}{4000} \sum_{i=1}^{30} x_i^2 - \prod_{i=1}^{30} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1,$$

onde $-600 \leq x_i \leq 600$, para $1 \leq i \leq 30$, e $\min(f_{11}) = f_{11}(0, \dots, 0) = 0$.

Família Shekel:

$$f(\mathbf{x}) = -\sum_{i=1}^m \left(\sum_{j=1}^4 (x_j - a_{i,j})^2 \right)^{-1},$$

para $0 \leq x_i \leq 10$, onde $1 \leq i \leq 4$. A função apresenta m mínimos locais nos pontos

$$\mathbf{x}_i \approx (a_{i,1}, a_{i,2}, a_{i,3}, a_{i,4})^T,$$

que assumem os valores $f(\mathbf{x}_i) \approx -\frac{1}{c_i}$. Para as funções f_{21} e f_{22} , m assume os valores 5 e 7, respectivamente. Os coeficientes c_i e $a_{i,j}$ são definidos conforme a tabela 1.

i	$a_{i,1}$	$a_{i,2}$	$a_{i,3}$	$a_{i,4}$	c_i
1	4	4	4	4	0,1
2	1	1	1	1	0,2
3	8	8	8	8	0,2
4	6	6	6	6	0,4
5	3	7	3	7	0,4
6	2	9	2	9	0,6
7	5	5	3	3	0,3

Tabela 1: Coeficientes de Shekel

Os resultados foram gerados para 20, 50 e 100 pólos iniciais, sendo cada experimento rodado 100 vezes, considerando um máximo de 500 fases históricas com duração de 100 gerações cada, passo inicial de 0,5, contradição mínima de 0,1 e efeito de crise máximo de 0,9. O limiar empregado foi de 0,001 para todas as funções, exceto para f_{21} e f_{22} , quando assume 9,5. As tabelas 2, 3 e 4 mostram os resultados de minimização obtidos. Os resultados são expressos usando média, desvio médio, mediana e moda dos valores mínimos encontrados, do número de iterações e do número de avaliações da função objetivo, além da taxa de sucesso, ou seja, do número de valores mínimos abaixo do limiar pelo número de experimentos para uma dada configuração.

De forma semelhante foram gerados resultados usando o algoritmo PSO, para populações de 20, 50 e 100 partículas iniciais, com $\Delta = 0,001$ e fator de inércia $w = 0,8$, sendo cada experimento rodado 30 vezes, considerando um máximo de 50000 gerações e as mesmas estatísticas utilizadas para o método dialético de otimização. As tabelas 5, 6 e 7 mostram os resultados de minimização obtidos.

	$m(0)$	\bar{y}	σ_y	$m_d(y)$
f_1	20	0,0006	0,0003	0,0007
	50	0,0004	0,0002	0,0004
	100	0,0004	0,0003	0,0004
f_5	20	0,0333	0,0515	0,0008
	50	0,3531	0,5918	0,0049
	100	0,0891	0,1145	0,0068
f_9	20	0,0007	0,0003	0,0009
	50	0,0008	0,0002	0,0009
	100	0,0007	0,0002	0,0009
f_{11}	20	0,0005	0,0002	0,0007
	50	0,0006	0,0002	0,0007
	100	0,0006	0,0002	0,0006
f_{21}	20	-9,8180	0,1716	-9,7787
	50	-9,8155	0,1900	-9,7673
	100	-9,7879	0,2005	-9,7244
f_{22}	20	-9,9142	0,2559	-9,8433
	50	-9,8217	0,2066	-9,7846
	100	-9,7806	0,2091	-9,7026

Tabela 2: Média, desvio médio e mediana dos valores mínimos (ODM)

	$m(0)$	\bar{k}	$m_d(k)$	$m_o(k)$	$\phi(\%)$
f_1	20	3864,36	2354	8	100
	50	2748,56	1402,5	3	100
	100	2422,17	4	2	100
f_5	20	31902,92	39584	50000	59
	50	38792,79	50000	50000	37
	100	40448,62	50000	50000	32
f_9	20	20533,91	11806	50000	79
	50	19988,28	8906	50000	88
	100	15427,92	4306	50000	94
f_{11}	20	3202,75	2106	3	100
	50	4340,83	4054	201	100
	100	3427,85	2901,5	201	100
f_{21}	20	1735,69	101	1	100
	50	2656,8	201,5	1	100
	100	2726,55	1	1	100
f_{22}	20	2124,12	101,5	1	100
	50	1816,81	1	1	100
	100	1121,98	1	1	100

Tabela 3: Número de iterações para minimização (ODM)

	$m(0)$	\bar{q}	$m_d(q)$	$m_o(q)$	$\phi(\%)$
f_1	20	9186,54	6508	160	100
	50	8527,84	7605	150	100
	100	9278,84	400	200	100
f_5	20	65605,84	80968	101800	59
	50	82385,58	104800	104800	37
	100	90697,24	109800	109800	32
f_9	20	42717,88	25412	101800	79
	50	44729,52	22612	104800	88
	100	40463,76	18412	109800	94
f_{11}	20	7838,66	6012	60	100
	50	13340,54	12908	5202	100
	100	16558,68	15603	10202	100
f_{21}	20	4398,56	2002	20	100
	50	8070,24	5203	50	100
	100	9334,88	100	100	100
f_{22}	20	5192,88	2003	20	100
	50	5534,9	50	50	100
	100	5446,6	100	100	100

Tabela 4: Número de avaliações da função (ODM)

	$m(0)$	\bar{y}	σ_y	$m_d(y)$
f_1	20	4800,5886	1159,2403	4884,2539
	50	2542,3720	926,2590	2217,9781
	100	1585,4484	571,9671	1428,3820
f_5	20	277,0581	72,2573	277,1713
	50	143,2292	33,5102	141,4566
	100	129,9445	50,2110	119,6252
f_9	20	167,4201	19,3534	161,2381
	50	130,1778	20,0383	131,6258
	100	116,3684	15,9510	115,5775
f_{11}	20	57,4452	15,6720	55,3680
	50	23,1885	8,0612	22,3225
	100	18,1383	7,1415	14,9742
f_{21}	20	-8,6186	1,7391	-9,6303
	50	-8,6305	1,8283	-9,6709
	100	-8,4603	1,8967	-9,6373
f_{22}	20	-8,6020	1,8185	-9,6293
	50	-9,1287	1,2324	-9,7488
	100	-9,6583	0,3574	-9,7707

Tabela 5: Média, desvio médio e mediana dos valores mínimos (PSO)

	$m(0)$	\bar{k}	$m_d(k)$	$m_o(k)$	$\phi(\%)$
f_1	20	50000	50000	50000	0
	50	50000	50000	50000	0
	100	50000	50000	50000	0
f_5	20	50000	50000	50000	0
	50	50000	50000	50000	0
	100	50000	50000	50000	0
f_9	20	50000	50000	50000	0
	50	50000	50000	50000	0
	100	50000	50000	50000	0
f_{11}	20	50000	50000	50000	0
	50	50000	50000	50000	0
	100	50000	50000	50000	0
f_{21}	20	10667,8	810	50000	80
	50	9196,86	1076	50000	83,33
	100	12304,3	903	50000	76,66
f_{22}	20	8735,3	406	50000	83,33
	50	7386,3	964	50000	86,66
	100	2045,13	248,5	1198	96,66

Table 6: Número de iterações para minimização (PSO)

	$m(0)$	\bar{q}	$m_d(q)$	$m_o(q)$	$\phi(\%)$
f_1	20	1000000	1000000	1000000	0
	50	2500000	2500000	2500000	0
	100	5000000	5000000	5000000	0
f_5	20	1000000	1000000	1000000	0
	50	2500000	2500000	2500000	0
	100	5000000	5000000	5000000	0
f_9	20	1000000	1000000	1000000	0
	50	2500000	2500000	2500000	0
	100	5000000	5000000	5000000	0
f_{11}	20	1000000	1000000	1000000	0
	50	2500000	2500000	2500000	0
	100	5000000	5000000	5000000	0
f_{21}	20	213356	16200	1000000	80
	50	459843,33	53800	2500000	83,33
	100	1230430	90300	5000000	76,66
f_{22}	20	174706	8120	1000000	83,33
	50	369315	48200	2500000	86,66
	100	204513,33	24850	119800	96,66

Tabela 7: Número de avaliações da função (PSO)

7. Discussão e Conclusões

Os resultados das tabelas 2, 3 e 4 mostram que o método de otimização proposto atingiu acertos de 100% para as funções f_1 , f_{11} , f_{21} e f_{22} , mostrando um excelente desempenho na otimização de uma função unimodal de alta dimensão (f_1), para uma função multimodal de alta dimensão e muitos mínimos locais f_{11} , e especialmente para funções multimodais de baixa dimensão mas com um pequeno número de mínimos locais fortemente atrativos (f_{21} e f_{22}), tendo atingido o mínimo absoluto dessas últimas usando apenas uma única iteração na maioria dos experimentos.

Os resultados para a minimização da função f_9 , uma função multimodal de alta dimensionalidade e muitos mínimos locais, podem ser classificados como entre bons e excelentes, mas ilustram que o aumento da quantidade de pólos iniciais pode ajudar no processo de otimização, já que incrementa a busca inicial. Já os resultados com a função f_5 mostram uma convergência mais lenta pois, apesar da baixa taxa de sucesso, os valores mínimos encontrados ficaram na vizinhança do mínimo absoluto.

Comparando-se os resultados expressos nas tabelas 2 e 4 com aqueles expressos nas tabelas 5 e 7, fica claro que, para as funções utilizadas neste trabalho, os resultados obtidos com o método dialético proposto são superiores tanto do ponto de vista da taxa de sucesso quanto levando em conta o número de avaliações da função objetivo, principalmente quando se consideram as funções f_1 e f_5 (unimodais de alta dimensão), f_9 e f_{11} (multimodais de alta dimensão e grande número de pontos de

mínimo), e f_{21} e f_{22} (multimodais de baixa dimensão e número pequeno de mínimos locais, mas fortemente atrativos): tanto para as funções unimodais quanto para as funções multimodais com muitos mínimos locais os resultados de taxa de sucesso para o algoritmo PSO foram muito ruins (0% de sucesso), enquanto os resultados para as funções multimodais com poucos mínimos locais e de baixa dimensionalidade ficaram entre 76% e 97% de sucesso, contra 100% de sucesso para o método dialético, mostrando que houve muitos resultados que convergiram para mínimos locais.

Esses resultados comprovam que é possível construir métodos de otimização buscando inspiração na Filosofia, e em especial em filosofias sistemáticas e investigativas como a dialética, com auxílio de elementos de Lógica Fuzzy e de Teoria da Informação, tais como o conceito de entropia.

É possível que a introdução do Princípio da Máxima Entropia em outros algoritmos de otimização consiga aumentar a capacidade exploratória dos algoritmos e, assim, torná-los mais robustos à influência de mínimos locais fortemente atrativos, no entanto é preciso que seja avaliada a possibilidade de gerar novas versões de outros algoritmos de busca e otimização.

8. Referências Bibliográficas

- [1] Cooper, D. E. *As Filosofias do Mundo: Uma Introdução Histórica*, Loyola, São Paulo, 2002.
- [2] Eberhart, R. & Kennedy, J. “A new optimizer using particle swarm theory”, *IEEE Symposium on Micro Machine and Human Science*, CIS-IEEE, Nagoya, Japan, 1995.
- [3] Eberhart, R. & Shi, Y. *Computational Intelligence: concepts to implementations*, Morgan Kaufmann, 2007.
- [4] Fan, J. L. & Ma, Y. L. “Some new fuzzy entropy formulas”, *Fuzzy Sets and Systems*, **128**: 277-284, 2002.
- [5] Goldmann, L. *Dialética e cultura*, Paz e Terra, Rio de Janeiro, 1967.
- [6] Kennedy, J. & Eberhart, R. “Particle swarm optimization”, *IEEE International Conference on Neural Networks*, CIS-IEEE, Perth, Australia, 1995.
- [7] Konder, L. *O futuro da Filosofia da Práxis: o pensamento de Marx no século XXI*, Paz e Terra, Rio de Janeiro, 1992.
- [8] Rahnamayan, S., Tizhoosh, H. R. & Salama, M. M. A. “A novel population initialization method for accelerating evolutionary algorithms”, *Computers and Mathematics with Applications*, (53): 1605-1614, 2007.
- [9] Yager, R. R. “On the entropies of fuzzy measures”, *IEEE Transactions on Fuzzy Systems*, **8**, (4): 453-461, 2000.