

3

Aplicação do Princípio Evolucionário em Otimização

Charles Robert Darwin nasceu em 12 de fevereiro de 1809 na cidade de Shrewsbury, Inglaterra. Após finalizar seus estudos em 1825 na *Shrewsbury School*, ingressou na universidade de Edimburgo, onde começou a estudar Medicina. Dois anos depois deixou a carreira e foi admitido na Universidade de Cambridge com o propósito de se converter em ministro da igreja Anglicana. Foi ali que conheceu o geólogo Adam Sedgwick e o naturalista John Stevens Henslow, com os quais realizou uma expedição científica ao redor do mundo após sua formatura em 1831. Foi nessa expedição que ele teve a oportunidade de observar diversas formações geológicas nas distintas ilhas e continentes além da ampla variedade de fósseis e organismos vivos.

Baseado nas suas observações geológicas, Darwin concluiu que a superfície da terra estava sendo submetida a constantes mudanças como resultado das forças naturais que atuavam de forma uniforme durante longos períodos de tempo. Entretanto, suas observações mais surpreendentes foram sobre a diversidade dos seres vivos, ao perceber nas ilhas Galápagos espécies de animais muito semelhantes às de outros lugares, mas que possuíam um tamanho diferente devido à sua estrutura e hábitos alimentares. Em função disto, ele deduziu que estas espécies não tinham surgido no local, mas que tinham migrado para as ilhas Galápagos, procedentes do continente onde, por não terem competidores e estarem isoladas geograficamente, sofreram uma rápida adaptação ao novo ambiente, resultando no aparecimento de novas espécies, sendo todas elas descendentes de um antepassado comum.

Em 1836, após seu retorno à Inglaterra, Darwin começou a aprimorar suas idéias sobre esta transmutação das espécies. Entretanto, a explicação da evolução dos organismos lhe surgiu após a leitura do livro “*Essay on the Principle of Population*” [36], publicado em 1798 pelo economista britânico Thomas Robert Malthus, que explicava como se mantinha o equilíbrio nas populações humanas. A tese básica de Malthus sustenta a existência de uma “desigualdade”, do tipo

“menor que”, entre a taxa de crescimento da população e a taxa de crescimento dos seus meios de subsistência. Darwin aplicou de imediato esta “desigualdade” de Malthus aos animais e plantas e, em 1859, publicou seu livro “*The Origin of Species*” [13]. Em essência, sua teoria defende que os membros jovens das distintas espécies competem intensamente por sua sobrevivência devido aos recursos limitados de subsistência disponíveis no seu meio ambiente. Os que sobrevivem tendem a incorporar variações naturais favoráveis ao processo de seleção natural, e estas variações se transmitem através da herança genética (estudada por Johann Gregor Mendel). Em consequência, cada geração tende a melhorar em termos de adaptação ao meio ambiente às gerações anteriores, e este processo gradual e contínuo é a causa da evolução das espécies. Embora a teoria de Darwin não é aceita, especialmente pelos segmentos religiosos, a comunidade científica reconheceu seu trabalho, dando a Darwin o título de membro da Academia Francesa de Ciências em 1878. Darwin morreu em 19 de abril de 1882 em Westminster.

Johann Gregor Mendel, nasceu em 22 de julho de 1822 na cidade de Heinzendorf, Republica Checa. Em 1843 ingressou no monastério de Santo Tomás de Brunn, em Áustria. Posteriormente, em 1851 ingressou na universidade de Viena, onde estudou história, botânica, física, química e matemáticas. A seu retorno ao monastério em 1854 iniciou uma série de trabalhos em plantas tentando descobrir o processo de transmissão das características hereditárias. Em 1866 publicou seu trabalho “*Versuche über Pflanzen-Hybride*” onde formulou as leis que levam seu nome e que constituem o fundamento de quase toda a genética moderna. Mendel morreu em 6 de Janeiro de 1884 na cidade de Brno, Áustria.

As leis de Mendel junto com o princípio da “sobrevivência dos mais aptos” de Darwin, explicam cientificamente o processo de evolução dos seres vivos. Entretanto, na atualidade o princípio Darwiniano, é considerado um caso especial da “sobrevivência dos estáveis” [37]. Isto verifica-se desde que o universo esta povoado por coisas estáveis, já que desde as estruturas moleculares até os conglomerados galácticos, são todas, em maior ou menor grau, padrões estáveis de átomos. Com isto o princípio Darwiniano, que governa a evolução dos seres vivos em ambientes orgânicos é um caso especial do “princípio evolucionário” que governa a evolução das entidades que formam parte dos diversos ambientes do cosmos.

Generalizando este raciocínio, John Holland aplicou com sucesso o princípio evolucionário em ambientes abstratos como é o caso do problemas de otimização. Seu trabalho, publicado em 1975, intitulado “*Adaptation in Natural and Artificial Systems*” [38], é considerado a semente de uma nova tecnologia de otimização denominada Algoritmos Genéticos (AGs).

Desta forma, o problema de otimização descrito no capítulo anterior pode ser resolvido através desta nova técnica. Nas próximas seções descreve-se uma versão computacional do princípio evolucionário em problemas de otimização de funções objetivo, e discutem-se os fundamentos teóricos que garantem o processo de convergência do AG.

3.1. Os Algoritmos Genéticos

Sem perda de generalidade, o problema de otimização será abordado num contexto de minimização da função de custo definida pela equação (41). Assim, o objetivo do AG é procurar o ponto θ^{\min} onde a função $\mathbf{h}(\theta) \geq 0, \forall \theta \in \mathfrak{R}^n$ possui seu mínimo valor global.

A interpretação computacional de Holland do princípio Darwiniano ao problema de otimização começa com a formulação de uma *população* inicial $\{\theta_t^{(1)}, \theta_t^{(2)}, \dots, \theta_t^{(N)}\}$ de N indivíduos. Cada indivíduo ou *cromossoma* $\theta_t^{(i)}$ desta população representa uma solução potencial do problema de minimização da função $\mathbf{h}(\theta)$. Durante cada iteração ou *geração* t do algoritmo, os *cromossomas* são *avaliados* usando alguma medida de *aptidão*. Uma nova geração de *cromossomas* são formadas selecionando os indivíduos mais aptos da população atual. Alguns membros da nova geração *evoluem* por meio dos operadores genéticos de *cruzamento* e *mutação* as quais são inspiradas nas leis Mendelianas da genética. Após várias gerações, é esperada a convergência do algoritmo para a solução ótima ou sub-ótima do problema.

Após esta interpretação computacional do princípio Darwiniano ao problema de otimização, Holland procedeu à codificação dos *cromossomas* $\theta_t^{(i)}$ numa representação binária $\theta_t^{(i)} = (b_1 b_2 \dots b_k \dots b_m)_t^{(i)}$ de m bits, onde $b_i \in \{0,1\}$ e o valor de m depende diretamente da precisão numérica adotada (maiores detalhes

de codificação e decodificação de *cromossomas* em [39,40]). Logo, este vetor binário tipicamente é armazenado numa estrutura computacional de *cadeias de caracteres*.

Sob estas considerações, a seguir mostra-se a estrutura do AG através do pseudo código denominado na literatura como *Programa Evolutivo* (PE) [39]:

Procedimento Programa Evolutivo

Início

$t \leftarrow 1$

Iniciar $\{\theta_t^{*(i)}\}_{i=1}^N$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$

Enquanto (condição de parada não é satisfeita) Fazer

Início

$t \leftarrow t + 1$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ **de** $\{\theta_{t-1}^{*(i)}\}_{i=1}^N$

Evoluir $\{\theta_t^{*(i)}\}_{i=1}^N$ **de** $\{\theta_t^{(i)}\}_{i=1}^N$

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$

Fim

Fim

onde $\{\theta_t^{*(i)}\}_{i=1}^N$ representa os indivíduos evoluídos a partir da população $\{\theta_t^{(i)}\}_{i=1}^N$, à exceção do estágio *Iniciar*, no qual representa a população inicial. A seguir detalha-se cada um dos procedimentos envolvidos no algoritmo PE.

Iniciar $\{\theta_t^{*(i)}\}_{i=1}^N$:

Neste primeiro estágio em $t = 1$, os *cromossomas* $\theta_t^{*(i)}$ da população representados por vetores binários, são inicializados aleatoriamente com valores $\{0,1\}$.

Avaliar $\{\theta_t^{*(i)}\}_{i=1}^N$:

Neste estágio, os *cromossomas* $\theta_t^{*(i)}$ da população são decodificados e avaliadas $\omega_t^{(i)} = \mathbf{h}(\theta_t^{*(i)})$ para obter a sua *aptidão normalizada* $\tilde{\omega}_t^{(i)}$ ao problema de minimização da função $\mathbf{h}(\theta)$. Sendo $\tilde{\omega}_t^{(i)}$ definido por:

$$\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{j=1}^N \omega_t^{(j)}} \quad (42)$$

Selecionar $\{\theta_t^{(i)}\}_{i=1}^N$ de $\{\theta_{t-1}^{*(i)}\}_{i=1}^N$:

Neste estágio, os indivíduos da população anterior $\{\theta_{t-1}^{*(i)}, \tilde{\omega}_{t-1}^{(i)}\}_{i=1}^N$ são selecionados, dando mais chance aos que possuem maior *aptidão normalizada*. Os indivíduos selecionados $\{\theta_t^{(i)}, N^{-1}\}_{i=1}^N$ formam a nova população. Note que a *aptidão normalizada* destes novos indivíduos $\theta_t^{(i)}$ são iguais a $\tilde{\omega}_t^{(i)} = N^{-1}$. Existem diversos critérios de seleção, entretanto o mais popular deles é o algoritmo da roleta [39,40].

Evoluir $\{\theta_t^{*(i)}\}_{i=1}^N$ de $\{\theta_t^{(i)}\}_{i=1}^N$

Neste estágio, alguns indivíduos da população atual sofrem alterações através dos operadores genéticos de *cruzamento* e *mutação*. Estas alterações são controladas através da probabilidade de *cruzamento* $\rho_c \in [0,1)$ e a probabilidade de *mutação* $\rho_m \in [0,1)$. Assim:

Para $i = 1, \dots, N/2$, $j = N - i + 1$, $u \sim \mathbf{U}(0,1)$, $k \sim \mathbf{U}(1..m - 1)$

$$(\theta_t^{(i)}, \theta_t^{(j)}) = \begin{cases} (\theta_t^{*(i)}, \theta_t^{*(j)}) & \text{se } u \leq \rho_c \\ (\theta_t^{(i)}, \theta_t^{(j)}) & \text{caso contrário} \end{cases} \quad (43)$$

Sendo que:

$$\begin{aligned} \theta_t^{*(i)} = (b_1 b_2 \dots b_k c_{k+1} \dots c_m) &\Leftrightarrow \theta_t^{(i)} = (b_1 b_2 \dots b_k b_{k+1} \dots b_m) \\ \theta_t^{*(j)} = (c_1 c_2 \dots c_k b_{k+1} \dots b_m) &\Leftrightarrow \theta_t^{(j)} = (c_1 c_2 \dots c_k c_{k+1} \dots c_m) \end{aligned} \quad (44)$$

Para $i = 1, \dots, N$, $u \sim \mathbf{U}(0,1)$, $k \sim \mathbf{U}(1..m - 1)$

$$\theta_t^{(i)} = \begin{cases} \theta_t^{*(i)} & \text{se } u \leq \rho_m \\ \theta_t^{(i)} & \text{caso contrário} \end{cases} \quad (45)$$

Sendo que:

$$\theta_t^{*(i)} = (b_1 b_2 \dots \bar{b}_k \dots b_m) \Leftrightarrow \theta_t^{(i)} = (b_1 b_2 \dots b_k \dots b_m) \quad (46)$$

onde \bar{b}_k é o complemento do valor binário b_k .

A operação genética de *cruzamento* definida pela equação (43), conhecida na literatura como *cruzamento* de um ponto, junto com a operação genética de

mutação definida pela equação (45), são os operadores genéticos padrões dos *cromossomas* com representação binária, embora a literatura apresente outras definições que potencialmente melhoram o desempenho do algoritmo em determinadas aplicações [39,40].

As evidências experimentais mostram que o mecanismo de busca descrito pelo algoritmo PE, adicionado à “regra de seleção elitista”, a qual seleciona e/ou preserva o melhor indivíduo evoluído pelo algoritmo a cada geração, converge para o ponto ótimo θ^{\min} , isto é, o ponto onde a função $h(\theta)$ possui seu mínimo valor global.

Apesar destas evidências empíricas mostrando a convergência dos AGs, eles têm sido muitas vezes criticados pela falta de um formalismo teórico que permita estudar suas propriedades de convergência. Neste sentido, na próxima seção são apresentados os fundamentos teóricos que sustentam os Algoritmos Genéticos.

3.2. Fundamentos Teóricos dos Algoritmos Genéticos

O próprio John Holland, ciente da falta de uma plataforma teórica dos AGs, desenvolveu estudos teóricos sobre a representação binária dos *cromossomas*, empregando a noção de *schema* que é um padrão genético que descreve um conjunto de *cromossomas* do espaço de busca com similaridades em certas posições. Tipicamente, o *schema* é construído introduzindo o símbolo *don't care* (*) no alfabeto {0,1} dos genes. Suponha-se, por exemplo, uma população de *cromossomas* representados por uma *cadeia de caracteres* de tamanho $m = 10$, logo, um possível *schema* desta população seria (*111100100) uma vez que descreve (ou representa) os *cromossomas* (1111100100) e (0111100100). Analogamente, o *schema* (*1*1100100) descreve as *cadeias de caracteres*: (0101100100), (1101100100), (0111100100) e (1111100100). De forma genérica, qualquer *schema* representa 2^r *cadeias de caracteres* onde r é o número de símbolos *don't care* (*). Além desta propriedade dos *schemas*, Holland definiu a *ordem* e o *comprimento* do *schema*. A ordem do *schema* S , denotada por $o(S)$, é o número de posições fixas de 0 e 1, presentes no *schema*. Por exemplo, os seguintes três *schemata* de tamanho $m = 10$, $S_1 = (**001*110)$,

$S_2 = (****00**0*)$, $S_3 = (11101**001)$, possuem as seguintes ordens: $o(S_1) = 6$, $o(S_2) = 3$ e $o(S_3) = 8$. Já o comprimento do *schema* S , denotado por $\delta(S)$, é a distância entre a primeira e última posição fixa da *cadeia de caracteres*. Por exemplo, $\delta(S_1) = 10 - 4 = 6$, $\delta(S_{21}) = 9 - 5 = 4$ e $\delta(S_3) = 10 - 1 = 9$.

Em função das propriedades dos *schemas*, Holland formulou seu conhecido *Teorema do Schema* que basicamente mostra que “*schemata* curtos e de baixa ordem tendem a se proliferar ou desaparecer nas gerações seguintes, de acordo com sua *aptidão* média” (veja [38] para maiores detalhes da demonstração). Embora este teorema ofereça um boa noção da maneira como os AGs funcionam, ele apresenta limitações por não explicar muitos aspectos complexos dos AGs e ao não apresentar medidas da sua convergência. Recentemente, esta teoria vem sendo estendida com resultados promissores [41-43].

Uma outra alternativa para fundamentar a convergência dos AGs, isto é, encontrar a solução θ^{\min} , onde a função $\mathbf{h}(\theta)$ possui seu mínimo valor global, foi desenvolvida para a representação binária dos *cromossomas*, usando a teoria de Markov em espaço de estados discretos [44-50]. Nesta análise, uma determinada população de indivíduos $\{\theta_t^{(1)}, \theta_t^{(2)}, \dots, \theta_t^{(N)}\}$ no tempo t , representa um *estado* que pertence ao espaço de estados \mathbf{B} . Assim, um determinado *estado* no tempo t é um vetor $\{(b_1 b_2 \dots b_k \dots b_m)_t^{(1)}, (b_1 b_2 \dots b_k \dots b_m)_t^{(2)}, \dots, (b_1 b_2 \dots b_k \dots b_m)_t^{(N)}\}$ de $m \cdot N$ bits. Em conseqüência \mathbf{B} é “discreto”, uma vez que possui um número “finito” de estados, os quais podem ser enumerados de 1 a n , sendo n a *cardinalidade* de \mathbf{B} . Sob esta visão, a trajetória ao longo do tempo da população de indivíduos (evoluídos pelo AG), pode ser descrita probabilisticamente como uma seqüência de estados finitos, a qual comumente é denominada de cadeia finita de Markov. Nesta cadeia, a probabilidade $p_{ij}(t)$ de transição do estado $i \in \mathbf{B}$ ao estado $j \in \mathbf{B}$ no tempo t é denominada *probabilidade de transição* de i para j em t . Se a *probabilidade de transição* é independente de t , então a cadeia é dita *homogênea*. Para uma determinada cadeia finita e homogênea, as *probabilidades de transição* podem ser agrupados numa *matriz de transição* $\mathbf{T} = (p_{ij})$. Em conseqüência, \mathbf{T} captura a dinâmica AG, pois descreve todas suas possíveis transições da população. Além disso, com o objetivo de analisar os diversos estágios do AG, a matriz \mathbf{T} pode ser decomposta num produto de matrizes $\mathbf{T} = \mathbf{C} \cdot \mathbf{M} \cdot \mathbf{S}$, de forma a capturar as

transições intermediárias causadas pelo *cruzamento* C , *mutação* M e *seleção* S , respectivamente. Para assegurar a convergência do AG, é necessário que T seja sempre *positiva* e *estocástica*. Uma matriz é dita *positiva* se todos seus elementos são maiores que 0, e é dita *estocástica* se a somatória de todos seus elementos é 1. Logo, sendo T *positiva* e *estocástica* garante-se que o AG possa visitar todos os estados de B . As condições impostas a T obrigam a matriz C ser *estocástica*, M ser *positiva* e *estocástica* e S ser *estocástica* com pelo menos um elemento positivo em cada coluna. Pode ser mostrado (maiores detalhes em [45]) que o operador genético de *cruzamento* definido pela equação (43) gera uma matriz C *estocástica*. Também é mostrado que o operador genético definido pela equação (45) gera uma matriz M *positiva* e *estocástica*. Entretanto, mostra-se também que o processo de seleção deve possuir a “regra de seleção elitista” antes ou depois deste estágio para garantir que a matriz C seja *estocástica* com pelo menos um elemento positivo em cada coluna. Em conclusão, esta fundamentação teórica dos AGs mostra matematicamente que o mecanismo de busca do algoritmo PE, acima descrito, adicionado da “regra de seleção elitista”, antes ou depois do estágio *Selecionar*, converge ao ponto global ótimo da função objetivo em análise.

Entretanto, a representação binária tradicional não é adequada para problemas numéricos multidimensionais de alta precisão [39] que utilizam a representação por números reais (ponto flutuante), para a qual a convergência dos AGs não foi ainda demonstrada. Neste contexto este trabalho propõe no próximo capítulo uma nova fundamentação teórica dos algoritmos genéticos baseados na representação em ponto flutuante dos *cromossomas* e na teoria de simulações estocásticas.

3.3. Resumo

Este capítulo teve como objetivo mostrar a aplicação do princípio evolucionário em problemas de otimização. Entretanto, a generalidade do princípio evolucionário permite aplicá-lo além do escopo de otimização, em problemas como o aprendizado dos MEE, assunto que será abordado no próximo capítulo.